UNIVERSIDADE FEDERAL DE SANTA CATARINA

DEPARTAMENTO DE MATEMÁTICA

ENTROPIA PARAMÉTRICA GENERALIZADA E A PROBABILIDADE DE ERRO

TESE SUBMETIDA À UNIVERSIDADE FEDERAL DE SANTA CATARINA PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS, ESPECIALIDADE EM MATEMÁTICA.

ALBERTINA ZATELLI

ABRIL - 1982

FLORIANÓPOLIS

SANTA CATARINA - BRASIL

Esta Tese foi julgada adequada para a obtenção do título de

"MESTRE EM CIÊNCIAS"

especialidade em Matemática, e aprovada em sua forma final pelo Curso de Pós-Graduação em Matemática da Universidade Federal de Santa Catarina

Prof. Inder Jeet Taneja, Ph.D.
Coordenador

BANCA EXAMINADORA:

Prof. Inder Jeet Taneja, Ph.D.
Orientador

Prof. Ruy Madsen Barbosa, Dr.

Prof. Gur Dial, Ph.D.

A G R A D E C I M E N T O S

A todos os amigos e colegas que direta ou indiretamente contribuiram para o bom termo do presente trabalho.

À Universidade Federal de Santa Catarina.

À Agnese,

Ida e

Isidoro

R E S U M O

Uma cota superior, bem como uma inferior, sobre a probabilidade de erro para o problema do reconhecimento de modelos é obtida em termos da entropia paramétrica generalizada, também chamada entropia de ordem α e grau β. A avaliação da cota superior requer o conhecimento da probabilidade "a priori" de funções densidade de probabilidade condicional. Também foi provado que as cotas aqui obtidas são mais precisas que as anteriormente conhecidas. Os casos particulares são a entropia quadrática, cota de Chernoff, distância Bayesiana, etc...; a cota inferior obtida é análoga à cota de Fano.

Também foram estudadas algumas propriedades novas da entropia paramétrica generalizada.

A B S T R A C T

An upper as well as lower bound on probability of error for the general pattern recognition problem is obtained in terms of the generalized parametric entropy called entropy of order α and degree β. Evaluation of the upper bound requires the knowledge of prior probabilities of the class-conditional probability density functions. It has also been proved that the bounds studied here are sharper than the previously known bounds. The particular cases are the quadratic entropy, Chernoff's bound Bayesian distance, while the lower bound is analogue to Fano's bound. Some new properties of the generalized parametric entropy have also been studied.

ÍNDICE

INTRODUÇÃO

Nos últimos anos foram desenvolvidas e estudadas extensivamente diversas medidas de informação (entropias), tais como a entropia de Shannon [20] e as entropias de ordem α, grau β, gama generalizada, entropia de ordem α e grau β, introduzidas por Renyi [18], Havdra e Charvat [12] e Daroczy [6], Arimoto [2], Sharma e Mittal [21] respectivamente. Algumas aplicações destas medidas se encontram no reconhecimento de modelos, estatística e teoria da codificação.

Neste trabalho estudaremos a entropia introduzida por Sharma e Mittal a qual chamaremos de entropia paramétrica generalizada (ou entropia de ordem α e grau β), apresentaremos algumas propriedades algébricas e analíticas interessantes desta medida, daremos cotas superiores e inferiores sobre a probabilidade de erro para um modelo de decisão bayesiano usando a entropia condicional paramétrica generalizada, sendo que a cota inferior obtida é uma generalização da cota de Fano.

CAPÍTULO I

Neste capítulo apresentaremos as notações e definições usadas no desenvolvimento deste estudo e resumiremos os resultados principais obtidos.

1.1. Notações

No desenvolvimento deste trabalho serão usadas as seguintes notações:

. Δ_n o conjunto das distribuições completas de n-probabilidades.

$$\Delta_n = \{P=(p_1,\ldots,p_n) / \sum_{i=1}^{n} p_i = 1,\ p_i \geq 0,\ \forall_i = 1,\ldots,n\} \tag{1.1}$$

. Δ_n^o o conjunto das distribuições completas de n-probabilidades todas não nulas.

$$\Delta_n^o = \{P=(p_1,\ldots,p_n) / \sum_{i=1}^{n} p_i = 1,\ p_i > 0,\ \forall_i = 1,\ldots,n\} \tag{1.2}$$

. P^o o vetor de n-probabilidades todas iguais.

$$P^o = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{1.3}$$

. P^i o vetor de n-probabilidades, no qual a i-ésima componente é 1 e as demais são nulas.

$$P^i = (0,0,\ldots,0,1,0,\ldots,0) \tag{1.4}$$

. p_M a maior componente do vetor $P=(p_1,p_2,\dots,p_n)$

$$p_M = \max_i \{p_1,p_2,\dots,p_n\}, \text{ para } P \in \Delta_n \quad (1.5)$$

. log indicará o logarítmo na base 2.

. ℓn indicará o logarítmo na base e.

Para as definições que se seguem consideraremos $X=(x_1,\dots,x_n)$ uma variável aleatória discreta com uma distribuição de probabilidade $P=(p_1,\dots,p_n)$, onde $p_i=p(x_i)$ e $P \in \Delta_n$.

1.2. Entropia de Shannon

Entropia como uma medida de informação, tem um papel chave na Teoria de Informação. O primeiro conceito de medida de incerteza foi introduzido por Shannon em 1948 [20], o qual passou a ser denominado de entropia de Shannon e é definida por:

$$H(X) = H(P) = H(p_1,\dots,p_n)) = -\sum_{i=1}^{n} p_i \log p_i \quad (1.6)$$

1.3. Entropia de Ordem α

Uma generalização da entropia de Shannon foi apresentada em 1961 por Renyi [18], [19] e denominada entropia de ordem α. Propriedades conhecidas desta entropia e sua caracterização foram recentemente estudadas por Azél e Daróczy [1].

A entropia de ordem α (ou entropia de Renyi) é definida por:

$$H_\alpha(X)=H_\alpha(P)=H_\alpha((p_1,\ldots,p_n))=\frac{1}{1-\alpha}\log\left[\sum_{i=1}^{n} p_i^\alpha\right],\ \alpha<0,\ \alpha\neq 1 \qquad (1.7)$$

Quando $\alpha \to 1$, a entropia de ordem α reduz-se a entropia de Shannon, isto é,

$$\lim_{\alpha\to 1} H_\alpha(X) = -\sum_{i=1}^{n} p_i \log p_i$$

1.4. Entropia de Grau β

Havdra e Charvat (1967) [12] e Daróczy (1970) [6], introduziram o conceito de entropia de grau β, que é definida por:

$$H^\beta(X)=H^\beta(P)=H^\beta((p_1,\ldots,p_n))=(2^{1-\beta}-1)^{-1}\left[\sum_{i=1}^{n} p_i-1\right],\ \beta>0,\beta\neq 1 \quad (1.8)$$

Novamente temos a entropia de Shannon como o limite de $H^\beta(X)$, quando $\beta \to 1$.

As propriedades e caracterização desta entropia foram estudadas por Taneja [22], [23], Aczél e Daróczy [1] e Mathie e Rathie [17].

1.5. Entropia Gama Generalizada

Arimoto (1971) [2], amplia a definição de entropia considerando-a como o infimo de uma forma especial de "imprecisão".

<u>DEFINIÇÃO</u>: Seja $f(x)$ uma função escalar com valores reais, definida e não-negativa em $(0,1]$, com derivadas contínuas em $(0,1]$ e tal que $f(1)=0$. Definimos funções de entropia generalizadas como

$$H_f(p_1,\ldots,p_n) = \inf \sum_{i=1}^{n} p_i \, f(\tilde{p}_i) \; , \tag{1.9}$$

onde a operação inf é tomada sobre todas as distribuições de probabilidade $\tilde{p}=(\tilde{p}_1,\tilde{p}_2,\ldots,\tilde{p}_n) \in \Delta_n$.

Arimoto considerou o seguinte exemplo:

$$f^{\gamma}(x) = \frac{1 - x^{1-\gamma}}{1-\gamma} \; , \; \gamma\neq 1, \; \gamma>0$$

$$f^{1}(x) = -\,\ell n \; x = \lim_{\gamma\to 1} f^{\gamma}(x).$$

Então segue que

$$_{\gamma}H(P) = \inf_{\tilde{p}} \sum_{i=1}^{n} p_i \, f^{\gamma}(\tilde{p}_i) =$$

$$\begin{cases} 1 - \max\limits_{i} p_i \; , \quad \text{para} \quad \gamma=0. & (1.10) \\[2ex] \dfrac{1 - \left[\sum\limits_{i=1}^{n} p_i^{1/\gamma} \right]^{\gamma}}{1-\gamma} \; , \; \text{para } \gamma\neq 1, \; \gamma>0 & (1.11) \\[2ex] -\sum\limits_{i=1}^{n} p_i \, \ell n \; p_i \; , \quad \text{para } \gamma=1 & (1.12) \end{cases}$$

Esta entropia é chamada entropia gama generalizada (ou entropia de Arimoto).

Propriedades importantes da entropia (1.11) foram estudadas por Boekee e Van Der Lubbe [4].

1.6. Entropia Paramétrica Generalizada

A parte original do nosso trabalho será basicamente desenvolvida para uma entropia introduzida por Sharma e Mittal (1975) |21|, denominada entropia paramétrica generalizada ou também chamada entropia de ordem α e grau β, a qual é definida por:

$$H_\alpha^\beta(X)=H_\alpha^\beta(P)=H_\alpha^\beta((p_1,\dots,p_n))=(2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}}-1\right], \quad (1.13)$$

$$\alpha,\beta>0;\ \alpha,\beta\neq 1$$

CASOS PARTICULARES

(i) Tomando $\alpha=\beta$ em (1.13) obtemos

$$H^\beta(P)=(2^{1-\beta}-1)^{-1}\left[\sum_{i=1}^{n} p_i^\beta - 1\right]$$

a qual é a entropia de grau β.

(ii) Tomando $2-\beta=\frac{1}{\gamma}=\gamma$ em (1.13) obtemos

$$_\gamma H(P)=(2^{\gamma-1}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{1/\gamma}\right)^\gamma-1\right].$$

a qual é a entropia gama generalizada.

(iii) No caso $\beta \to 1$, a entropia paramétrica generalizada reduz-se simplesmente à entropia de ordem α

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left[\sum_{i=1}^{n} p_i^\alpha \right].$$

(iv) Teremos a entropia de Shannon quando $\beta \to 1$ e $\alpha \to 1$

$$H(P) = - \sum_{i=1}^{n} p_i \log p_i \; .$$

1.7. Regra da Decisão e Probabilidade de Erro no Reconhecimento de Modelos

Consideremos o problema da teoria da decisão, de classificar uma observação X como vinda de uma de n possíveis classes (hipóteses) $C=(c_1,\ldots,c_n)$. Denotemos $P_i=P_r\{C=c_i\}$, $i=1,2,\ldots,n$ "a priori" probabilidade das classes e denotemos $f_1(x),f_2(x),\ldots$ $\ldots,f_n(x)$ a função densidade condicional dada a verdadeira classe ou hipótese, isto é, $f_i(x)=P_r\{X=x/C=c_i\}$, $i=1,2,\ldots,n$. Suponhamos que $f_i(x)$ são completamente conhecidas. Dada qualquer observação $X=x$, podemos calcular "a posteriori" probabilidade condicional de C pela regra de Bayes.

$$P(c_j/x)=P_r\{C=c_j/X=x\} = \frac{P_j \; f_j(x)}{\sum_{i=1}^{n} P_i \; f_i(x)} \; , \; j=1,\ldots,n$$

É bem conhecido [9] (Ferguson (1967)) que a regra de decisão que minimiza a probabilidade de erro é a regra de decisão de Bayes a qual escolhe a hipótese com a maior probabilidade posterior. Usando esta regra, a probabilidade de erro para um dado X=x é expressa por:

$$P(e/x) = 1 - \max_i \{P(c_1/x), P(c_2/x), \ldots, P(c_n/x)\}$$

Antes de observar X, a probabilidade de erro P(e) associada à X é definida como a probabilidade de erro esperada após observá-lo, isto é,

$$P(e) = E_x\left[1 - \max_i \{P(c_1/x), \ldots, P(c_n/x)\}\right]$$

$$= 1 - E_x\left[\max_i \{P(c_1/x), \ldots, P(c_n/x)\}\right]$$

(i) Em [13], [10], [14], são obtidas cotas superiores sobre a probabilidade de erro em termos de entropia de Shannon

$$P(e) \leqslant \frac{1}{2} I, \tag{1.14}$$

onde I é a expectativa dada por:

$$I = E_x\left[H(C/x)\right]. \tag{1.15}$$

H(C/x) é a bem conhecida entropia condicional de Shannon de C dado X=x e é dada por:

$$H(C/x) = -\sum_{i=1}^{n} P(c_i/x) \log P(c_i/x) \tag{1.16}$$

A cota em (1.14) é baseada na seguinte desigualdade:

$$1 - p_M \leqslant \frac{-1}{2} \sum_{i=1}^{n} p_i \log p_i = \frac{1}{2} H(P) \tag{1.17}$$

onde $p_M = \max_i \{p_1,\ldots,p_n\}$ e $P=(p_1,\ldots,p_n) \in \Delta_n$

sendo $\Delta_n = \{P = (p_1,\ldots,p_n) \mid p_i \geqslant 0, \sum_{i=1}^{n} p_i = 1\}$,

o conjunto de todas as distribuições completas de n-probabilidades.

(ii) Em [3], uma cota superior sobre a probabilidade de erro foi obtida em termos de entropia de Rényi (entropia de ordem α) e é dada por:

$$P(e) \leqslant \frac{1}{2} I_2 \tag{1.18}$$

onde $$I_\alpha = E_x \left[H_\alpha(C/x)\right] \tag{1.19}$$

e $$H_\alpha(C/x) = \frac{1}{1-\alpha} \log \left[\sum_{i=1}^{n} P(c_i/x)^\alpha\right], \ \alpha > 0, \ \alpha \neq 1 \tag{1.20}$$

A cota (1.18) é mais próxima da igualdade que (1.14) desde que H_α é uma função decrescente de α. Se assumirmos que $P(e/x) \geqslant \frac{1}{2}$ então é obtida uma cota mais geral dada por:

$$P(e) \leqslant \frac{1}{2} I_\alpha \quad \text{para todo } \alpha > 0 \tag{1.21}$$

enquanto que para o caso de uma distribuição completa de duas probabilidades temos

$$P(e/x) \leqslant \frac{1}{2} I_\alpha \quad \text{para todo } x.$$

Foi também obtida uma cota inferior sobre a probabilidade de erro em em termos de entropia de ordem α [3] a qual é análoga à cota de Fano.

Os resultados ligados com a entropia gama generalizada e probabilidade de erro foram estudados por Arimoto [2] e Taneja [26].

Os resultados acima, concernentes à entropia de ordem α serão discutidos em detalhes no capítulo II. Enquanto que no capítulo III estudaremos algumas das novas propriedades da entropia paramétrica generalizada (ou entropia de ordem α e grau β) caracterizada por Sharma e Mittal [21]. Uma cota superior bem como uma cota inferior sobre a probabilidade de erro para o problema do reconhecimento de modelos é obtido em termos da medida paramétrica generalizada.

CAPÍTULO II

ENTROPIA DE ORDEM α E A PROBABILIDADE DE ERRO

Neste capítulo vamos estudar a entropia de ordem α introduzindo algumas novas propriedades, determinaremos cotas superiores e inferiores sobre a probabilidade de erro em termos da entropia condicional de ordem α e daremos uma generalização da Cota de Fano.

§2.1 - Propriedades da Entropia de Ordem α

A entropia de ordem α (ou entropia de Renyi) [18], [19]; já definida no capítulo I é dada por:

$$H_\alpha(P) = \frac{1}{1-\alpha} \log \left[\sum_{i=1}^{n} p_i^\alpha \right], \quad \alpha \geqslant 0, \ \alpha \neq 1 \tag{2.1}$$

para todo $P \in \Delta_n$ onde $\Delta_n = \{P=(p_1,\ldots,p_n);\ p_i \geqslant 0;\ \sum_{i=1}^{n} p_i = 1\}$ é o conjunto de todas as distribuições completas de n-probabilidades.

Observação: O conjunto de todas as distribuições completas de n-probabilidades não nulas será denotado por:

$$\Delta_n^o = \{P=(p_1,\ldots,p_n);\ p_i > 0;\ \sum_{i=1}^{n} p_i = 1\}$$

Os seguintes dois lemas foram provados em [18].

LEMA 2.1 - $H_\alpha(P)$ é uma função contínua e decrescente de α.

LEMA 2.2 - $H_o(P) = \log n$, para todo $P \varepsilon \Delta_n$. (2.2)

$$H_\alpha(P^o) = - \log \frac{1}{n} \text{, para todo } \alpha > 0 \tag{2.3}$$

$$H_\alpha(P^i) = 0, \quad \text{para todo } \alpha > 0,\ i=1,2,\ldots,n \tag{2.4}$$

$$\lim_{\alpha\to\infty} H_\alpha(P) = H_\infty(P) = -\log p_M, \quad \text{para todo } P \varepsilon \Delta_n \tag{2.5}$$

LEMA 2.3 - Seja $p_M = \max\{p_1,\ldots,p_n\}$, $\alpha > 0$, $\alpha \neq 1$ e $P \varepsilon \Delta_n$, então:

$$H_\alpha(1-p_M,p_M) \leqslant H_\alpha(P) \tag{2.6}$$

PROVA - Seja $P \varepsilon \Delta_n$

Sem perda de generalidade podemos supor $p_M = p_n$, onde $p_M = \max\{p_1,\ldots,p_n\}$.

1º Caso: $0 < \alpha < 1$,

Sabemos que:

$$\sum_{i=1}^{n} p_i^\alpha \geqslant \left(\sum_{i=1}^{n} p_i\right)^\alpha, \quad \alpha < 1 \tag{2.7}$$

então:

$$\sum_{i-1}^{n} p_i^{\alpha} = \sum_{i=1}^{n-1} p_i^{\alpha} + p_n^{\alpha} \geqslant (\sum_{i=1}^{n-1} p_i)^{\alpha} + p_n^{\alpha} = (1-p_M)^{\alpha} + p_M^{\alpha} \tag{2.8}$$

Desde que log é uma função crescente temos que:

$$\log \left[\sum_{i=1}^{n} p_i^{\alpha}\right] \geqslant \log\left[(1-p_M)^{\alpha} + p_M^{\alpha}\right] . \tag{2.9}$$

Multiplicando a desigualdade acima pelo fator positivo $\frac{1}{1-\alpha}$, pois $0 < \alpha < 1$ obtemos:

$$\frac{1}{1-\alpha} \log\left[\sum_{i=1}^{n} p_i^{\alpha}\right] \geqslant \frac{1}{1-\alpha} \log\left[(1-p_M)^{\alpha} + p_M^{\alpha}\right]$$

ou seja:

$$H_{\alpha}(1-p_M,p_M) \leqslant H_{\alpha}(P),\ P \in \Delta_n,\ 0 < \alpha < 1.$$

<u>2º Caso</u>: $\alpha > 1$,

neste caso as desigualdades (2.7), (2.8) e (2.9) ficam invertidas; desde que para $\alpha > 1$, $\frac{1}{1-\alpha}$ é um fator negativo obtemos novamente a desigualdade desejada e assim completamos a demonstração do lema.

<u>TEOREMA 2.1</u> - Seja $P \in \Delta_n$, então:

a) $H_{\alpha}(P)$ é uma função estritamente côncava com com respeito à P, para todo $0 < \alpha \leqslant 1$.

b) Para o caso $n=2$, $H_\alpha(P)$ é estritamente côncava com respeito à P, para $0 < \alpha \leqslant 2$.

c) Para todo $\alpha > 1$, existe um $n' > 2$ tal que $H_\alpha(P)$ não é côncava nem convexa com respeito à P, para todo $n > n'$.

d) Para todo $\alpha > 2$ e $n \geqslant 2$, $H_\alpha(P)$ não é côncava nem convexa, com respeito à P.

PROVA:

a) $0 < \alpha \leqslant 1, \quad n \geqslant 2$

Para $\alpha=1$, $H_\alpha(P)$ é a entropia de Shannon, que é bem conhecido ser estritamente côncava.

Suponhamos então $0 < \alpha < 1$:

Sejam P e $Q \in \Delta_n$, $P \neq Q$, e seja λ um número tal que $0 \leqslant \lambda \leqslant 1$. Visto que $w(x)=x^\alpha$ é uma função estritamente côncava com respeito à x, para $0 < \alpha < 1$, temos:

$$(\lambda p_i + (1-\lambda)q_i)^\alpha > \lambda p_i^\alpha + (1-\lambda)q_i^\alpha, \quad i=1,2,\ldots,n$$

portanto:

$$\sum_{i=1}^{n} (\lambda p_i+(1-\lambda)q_i)^\alpha > \sum_{i=1}^{n} (\lambda p_i^\alpha+(1-\lambda)q_i)=\lambda \sum_{i=1}^{n} p_i^\alpha+(1-\lambda) \sum_{i=1}^{n} q_i^\alpha$$

Desde que log é uma função estritamente crescente obtemos:

$$\log\left[\sum_{i=1}^{n}(\lambda p_i+(1-\lambda)q_i)^{\alpha}\right]>\log\left[\lambda\sum_{i=1}^{n}p_i^{\alpha}+(1-\lambda)\sum_{i=1}^{n}q_i^{\alpha}\right]$$

$$>\lambda\log\left[\sum_{i=1}^{n}p_i^{\alpha}\right]+(1-\lambda)\log\left[\sum_{i=1}^{n}q_i^{\alpha}\right] \qquad (2.10)$$

sendo que a relação (2.10) deve-se ao fato de log ser uma função côncava; assim, multiplicando ambos os lados da desigualdade acima, pelo valor positivo $\frac{1}{1-\alpha}$ $(0 < \alpha < 1)$, obtemos:

$$\frac{1}{1-\alpha}\log\left[\sum_{i=1}^{n}(\lambda p_i+(1-\lambda)q_i)^{\alpha}\right]>\frac{\lambda}{1-\alpha}\log\sum_{i=1}^{n}p_i^{\alpha}+\frac{(1-\lambda)}{1-\alpha}\log\sum_{i=1}^{n}q_i^{\alpha} \quad , (2.11)$$

ou seja:

$$H_{\alpha}(\lambda P + (1-\lambda Q) > \lambda H_{\alpha}(P)+(1-\lambda)H_{\alpha}(Q)$$

e isto demonstra (a).

b) <u>$1 < \alpha \leqslant 2;\quad n = 2$</u>

Sabemos que:

$$H_{\alpha}(p,1-p) = \frac{1}{1-\alpha}\log(p^{\alpha}+(1-p)^{\alpha}),\ p\ \varepsilon\ [0,1]$$

Portanto a primeira derivada de $H_{\alpha}(p,1-p)$ é dada por:

$$\frac{dH_{\alpha}(p,1-p)}{dp}=\frac{1}{1-\alpha}\,\frac{\alpha\,p^{\alpha-1}-\alpha(1-p)^{\alpha-1}}{(\ell n\ 2)(p^{\alpha}+(1-p)^{\alpha})}=\frac{\alpha}{(\ell n\ 2)(1-\alpha)}\cdot\frac{p^{\alpha-1}-(1-p)^{\alpha-1}}{(p^{\alpha}+(1-p)^{\alpha})}$$

fazendo q=1-p, temos:

$$\frac{d\,H_\alpha(p,1-p)}{dp} = \frac{\alpha}{(\ell n\ 2)(1-\alpha)} \cdot \frac{p^{\alpha-1}-q^{\alpha-1}}{p^\alpha + q^\alpha} \tag{2.12}$$

$$\frac{d^2H_\alpha(p,1-p)}{dp^2} = \frac{\alpha}{1-\alpha} \cdot \frac{(p^\alpha+ q^\alpha)^{-2}}{\ell n\ 2} \left\{ (p^\alpha + q^\alpha)\left[(\alpha-1)p^{\alpha-2}+(\alpha-1)q^{\alpha-2}\right] - \right.$$

$$\left. - (p^{\alpha-1}-q^{\alpha-1})\left[\alpha p^{\alpha-1}-\alpha q^{\alpha-1}\right]\right\} \tag{2.13}$$

$$\frac{d^2H_\alpha(p,1-p)}{dp^2} = \frac{\alpha}{1-\alpha}\ \frac{(p^\alpha+q^\alpha)^{-2}}{(\ell n\ 2)} \left[\alpha(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})\right.$$

$$\left. - \alpha(p^{\alpha-1}-q^{\alpha-1})^2\right]$$

$$= \frac{\alpha}{1-\alpha}\ \frac{(p^\alpha+q^\alpha)^{-2}}{(\ell n\ 2)} \left[\alpha(\cancel{p^{2\alpha-2}}+p^\alpha q^{\alpha-2}+p^{\alpha-2}q^\alpha+\cancel{q^{2\alpha-2}}-\cancel{p^{2\alpha-2}} + \right.$$

$$\left. + 2p^{\alpha-1}q^{\alpha-1}-\cancel{q^{2\alpha-2}})-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})\right]$$

$$= \frac{\alpha(p^\alpha+q^\alpha)^{-2}}{(1-\alpha)(\ell n\ 2)} \left[\alpha p^{\alpha-2}q^{\alpha-2}\underbrace{(p^2+2pq+q^2)}_{=1}-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})\right]$$

$$= \frac{\alpha(p^\alpha+q^\alpha)^{-2}}{(1-\alpha)(\ell n\ 2)} \left[\alpha p^{\alpha-2}q^{\alpha-2}-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})\right]$$

Assim, a segunda derivada de $H_\alpha(p,p-1)$ é dada por:

$$\frac{d^2H_\alpha(p,1-p)}{dp^2} = \frac{\alpha(p^\alpha+q^\alpha)^{-2}}{(1-\alpha)(\ell n\ 2)}\left[\alpha p^{\alpha-2}q^{\alpha-2}-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})\right] \qquad (2.14)$$

onde $q=1-p$. Denotando o fator do colchete direito por $h_\alpha(p)$, isto é, denotando

$$h_\alpha(p)=\alpha p^{\alpha-2}q^{\alpha-2}-(p^\alpha+q^\alpha)(p^{\alpha-2}+q^{\alpha-2})$$

temos:

$$h_\alpha(p)=(p^{\alpha-2}q^{\alpha-2})\left[\alpha\ -(p^\alpha+q^\alpha)(p^{2-\alpha}+q^{2-\alpha})\right]$$

então:

$$(p^{\alpha-2}q^{\alpha-2})^{-1}\ h_\alpha(p) = \alpha\ -(p^\alpha+q^\alpha)(p^{2-\alpha}+q^{2-\alpha}) \qquad (2.15)$$

$$\geqslant \alpha - (\tfrac{1}{2})^{1-\alpha} \qquad (2.16)$$

$$= \alpha - 2^{\alpha-1} > 0, \qquad (2.17)$$

para todo $1 < \alpha \leqslant 2$.

De fato: Seja $q = 1 - p$ e $p \in (0,1)$.

Desde que $1=(p+q)^2 > p^2+q^2$, para $\alpha=2$, temos:

$$(p^{2-2}q^{2-2})^{-1}\ h_2(P)=2-(p^2+q^2)(p^{2-2}+q^{2-2})=2-(p^2+q^2)2$$

$$=2(1-(p^2+q^2)) > 0=2-(\tfrac{1}{2})^{1-2} \quad .$$

Isto prova (2.16) e (2.17) para $\alpha = 2$.

Suponhamos agora $1 < \alpha < 2$. A seguinte desigualdade é válida:

$$(p^\alpha+q^\alpha) \leqslant (p+q)^\alpha = 1, \quad \text{por} \quad (2.7)$$

multiplicando a desigualdade acima pelo fator positivo $(p^{2-\alpha}+q^{2-\alpha})$ obtemos:

$$(p^\alpha+q^\alpha)(p^{2-\alpha}+q^{2-\alpha}) < p^{2-\alpha}+q^{2-\alpha} \qquad (2.18)$$

consideremos a função

$$u(p) = p^{2-\alpha}+q^{2-\alpha}, \quad \text{onde} \quad q=1-p$$

com

$$u'(p) = (2-\alpha)(p^{1-\alpha}-q^{1-\alpha})$$

e

$$u''(p) = (2-\alpha)(1-\alpha)(p^{-\alpha}+q^{-\alpha}),$$

de modo que $u''(p) < 0$ para todo $0 < p < 1$, assim, $u(p)$ é uma função estritamente côncava, e assume seu único máximo em $u'(p) = 0$, isto é, $p = \frac{1}{2}$. Portanto, (2.18) é menor ou igual à

$$\left(\frac{1}{2}\right)^{2-\alpha} + \left(\frac{1}{2}\right)^{2-\alpha} = 2\left(\frac{1}{2}\right)^{2-\alpha} = \left(\frac{1}{2}\right)^{1-\alpha}$$

Concluímos assim que para $1 < \alpha \leqslant 2$

$$(p^\alpha+q^\alpha)(p^{2-\alpha} + q^{2-\alpha}) < \left(\frac{1}{2}\right)^{1-\alpha}, \qquad (2.19)$$

multiplicando (2.19) por (-1) e somando α à ambos os membros da desigualdade assim obtida temos:

$$\alpha-(p^{\alpha}+q^{\alpha})(p^{2-\alpha}+q^{2-\alpha}) \geqslant \alpha-(\frac{1}{2})^{1-\alpha} = \alpha-2^{\alpha-1}$$

o que demonstra (2.16).

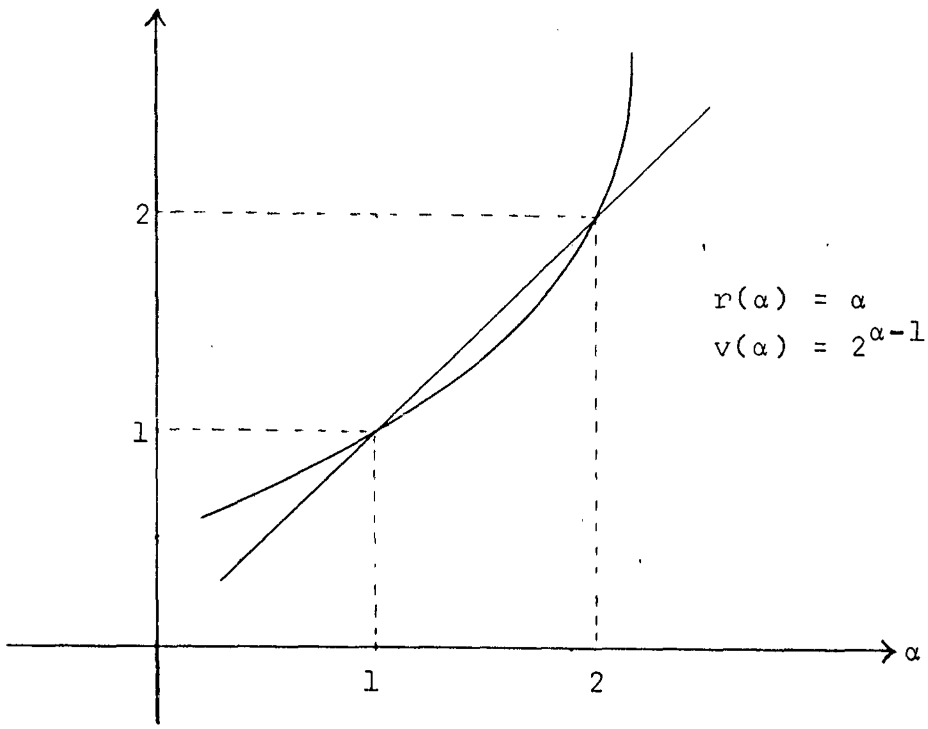
Consideremos as funções:

$$r(\alpha) = \alpha \qquad r(1) = v(1) = 1$$

$$v(\alpha) = 2^{\alpha-1} \qquad r(2) = v(2) = 2$$

Como o gráfico de $r(\alpha)$ consiste de uma reta que passa pelos pontos (1,1) e (2,2), os quais pertencem também ao gráfico de $v(\alpha)$ e sendo $v(\alpha)$ uma função estritamente convexa com respeito à α, temos (2.17), ou seja:

$$\alpha-2^{1-\alpha} > 0, \quad \text{para} \quad 1 < \alpha < 2.$$



Assim (2.15), (2.16) e (2.17) implicam que para $1 < \alpha \leq 2$, a segunda derivada de $H_\alpha(p,1-p)$ é negativa, deste modo provando (b).

A definição abaixo será usada na demonstração do item (c).

<u>DEFINIÇÃO 2.1</u> - $H_\alpha(P)$ é uma função côncava com respeito à $P \in \Delta_n$, se e somente se, para todo $P,Q \in \Delta_n$,

$$(\nabla H_\alpha(P) - \nabla H_\alpha(Q))(P-Q) \leq 0 \qquad (2.20)$$

c) <u>$\alpha > 1, \quad n > 2$:</u>

Desde que: $$\frac{\partial H_\alpha(P)}{\partial p_i} = \frac{\alpha}{1-\alpha} \cdot \frac{p_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^\alpha} \qquad (2.21)$$

Segue-se que

$$(\nabla H_\alpha(P) - \nabla H_\alpha(Q))(P-Q) =$$

$$= \frac{\alpha}{1-\alpha} \sum_{i=1}^{n} \left[\left(\frac{p_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^\alpha} - \frac{q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^\alpha} \right) (p_i - q_i) \right] =$$

$$= \frac{\alpha}{1-\alpha} \sum_{i=1}^{n} \left[\frac{p_i^\alpha - p_i^{\alpha-1} q_i}{\sum_{i=1}^{n} p_i^\alpha} - \frac{q_i^{\alpha-1} p_i - q_i^\alpha}{\sum_{i=1}^{n} q_i^\alpha} \right] =$$

$$= \frac{\alpha}{1-\alpha}\left[1 - \frac{\sum_{i=1}^{n} p_i^{\alpha-1} q_i}{\sum_{i=1}^{n} p_i^{\alpha}} - \frac{\sum_{i=1}^{n} p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}} + 1\right] =$$

$$= \frac{\alpha}{1-\alpha}\left[2 - \left(\frac{\sum_{i=1}^{n} p_i^{\alpha-1} q_i}{\sum_{i=1}^{n} p_i^{\alpha}} + \frac{\sum_{i=1}^{n} p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}}\right)\right] =$$

Assim para $\alpha > 1$, $\frac{\alpha}{1-\alpha} < 0$ e $H_\alpha(P)$ é côncava, se e somente se,

$$\frac{\sum_{i=1}^{n} q_i p_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}} + \frac{\sum_{i=1}^{n} p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}} < 2. \tag{2.22}$$

O seguinte exemplo estabelece a existência de n para o qual (2.22) não é válida. Sejam P e Q tais que $p_1=1$, $p_i=0$, $i\neq 1$; $q_1=a$, $q_i=\frac{1-a}{n-1}$, $i\neq 1$, $i=1,2,\ldots,n$, então

$$\frac{\sum_{i=1}^{n} p_i q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}} = \frac{a^{\alpha-1}}{a^{\alpha} + (n-1)\left(\frac{1-a}{n-1}\right)^{\alpha}} = \frac{a^{\alpha-1}}{a^{\alpha} + (1-a)^{\alpha}(n-1)^{1-\alpha}} \tag{2.23}$$

Para $\alpha > 1$, seja $a = (n-1)^{(1-\alpha)/\alpha}$ de modo que (2.23) é igualada à

$$\frac{(n-1)^{(\alpha-1)(1-\alpha)/\alpha}}{(n-1)^{\alpha(1-\alpha)/\alpha} + (1-a)^{\alpha}(n-1)^{(1-\alpha)}} = \frac{(n-1)^{(\alpha-1)(1-\alpha)/\alpha}}{(n-1)^{(1-\alpha)}\ (1+(1-a)^{\alpha})} \qquad (2.24)$$

Desde que $0 \leq a \leq 1$, $0 \leq (1-a)^{\alpha} \leq 1$, (2.24) é maior ou igual à

$$\frac{(n-1)^{(\alpha-1)(1-\alpha)/\alpha}}{2(n-1)^{(1-\alpha)}} = 1/2 \quad (n-1)^{(\alpha-1)(1-\alpha)(\alpha+(\alpha-1)\alpha/\alpha} =$$

$$= \frac{1}{2}\ (n-1)^{(\alpha-1)[(1-\alpha)+\alpha]/\alpha} = \frac{1}{2}\ (n-1)^{(\alpha-1)/\alpha} \qquad (2.25)$$

Assim se tomarmos $(n-1) > 4^{\alpha/(\alpha-1)}$, isto é, $n > 4^{\alpha(\alpha-1)}+1$, teremos:

$$\frac{\sum_{i=1}^{n} p_i\ q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}} > 2, \text{ o que invalida (2.22)}$$

De fato:

$$\frac{\sum_{i=1}^{n} q_i\ p_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}} = a, \quad \text{logo}$$

$$\frac{\sum_{i=1}^{n} q_i\ p_i^{\alpha-1}}{\Sigma p_i} + \frac{\sum_{i=1}^{n} p_i\ q_i^{\alpha-1}}{\sum_{i=1}^{n} q_i^{\alpha}} > 2 + a > 2 .$$

d) $\underline{\alpha > 2,\ n = 2 \ \ e\ \varepsilon > 0}$

Suficientemente pequeno, temos:

$$\frac{d^2H_\alpha(\frac{1}{2}, \frac{1}{2})}{dp^2} = \frac{-4\alpha}{\ell n2} < 0 \tag{2.26}$$

$$\frac{d^2H_\alpha(\varepsilon, 1-\varepsilon)}{dp^2} = \frac{d^2H_\alpha(1-\varepsilon, \varepsilon)}{dp^2} > 0 \tag{2.27}$$

De fato: De acordo com (2.14) a segunda derivada de $H_\alpha(\frac{1}{2}, \frac{1}{2})$ é da da por:

$$\frac{d^2H_\alpha(\frac{1}{2}, \frac{1}{2})}{dp^2} = \frac{\alpha}{(1-\alpha)} \cdot \frac{\left[(\frac{1}{2})^\alpha + (\frac{1}{2})^\alpha\right]^{-2}}{\ell n2} \left\{ \alpha(\frac{1}{2})^{\alpha-2} . (\frac{1}{2})^{\alpha-2} - \right.$$

$$\left. - \left[(\frac{1}{2})^\alpha + (\frac{1}{2})^\alpha\right] \left[(\frac{1}{2})^{\alpha-2} + (\frac{1}{2})^{\alpha-2}\right] \right\}$$

$$= \frac{\alpha}{(1-\alpha)} \frac{\left[2\ 2^{-\alpha}\right]^{-2}}{\ell n2} \left\{ \alpha\, 2^{(2-\alpha)2} - (2.2^{-\alpha})(2.2^{2-\alpha}) \right\}$$

$$= \frac{\alpha}{(1-\alpha)\ell n2} \left\{ 2^{(1-\alpha)(-2)} \quad 2^{(2-\alpha)2} \quad [\alpha-1] \right\}$$

$$= \frac{-\alpha}{\ell n 2}\, 2^{2\alpha-2+4-2\alpha} = \frac{-\alpha}{\ell n 2} \cdot 2^2$$

portanto: $$\frac{d^2 H_\alpha(\frac{1}{2}, \frac{1}{2})}{dp^2} = \frac{-4\alpha}{\ell n 2} \qquad (2.28)$$

As relações (2.26) e (2.27) implicam que para $\alpha > 2$, $n=2$, $P=(p,1-p)$ $H_2(P)$ não é côncava nem convexa. Desde que $P=(p,1-p)\ \varepsilon\ \Delta_2$ pode ser extendido para $P'=(p,1-p,0,\ldots,0)\varepsilon\Delta_n$, $n\geq 2$ adicionando zeros e $H_2(P)=H_2(P')$, portanto, não concavidade (não convexidade) em Δ_2 implica em não concavidade (não convexidade) em em Δ_n, para $n > 2$ e $\alpha > 2$.

<u>RESUMINDO</u>:

(i) Para $\alpha \leq 1$, $H_\alpha(P)$ é côncava para todo $n \geq 2$.

(ii) Para $1 < \alpha \leq 2$ existe algum "n" para o qual $H_\alpha(P)$ é côncava, e existe algum "n" para o qual não é.

(iii) Para $\alpha > 2$ $H_\alpha(P)$ não é côncava nem convexa para todo $n \geq 2$.

<u>DEFINIÇÃO 2.2</u>: Uma função numérica θ definida sobre um conjunto convexo $\Gamma\ \varepsilon\ \mathbb{R}^n$ é pseudocôncava em Γ ([15], [16]), se para x^1, $x^2\ \varepsilon\ \Gamma$:

$$\nabla\theta(x^1)(x^2-x^1) \leq 0 \text{ implica } \theta(x^2) \leq \theta(x^1) \qquad (2.29)$$

<u>TEOREMA 2.2</u> - $H_\alpha(P)$ é pseudocôncava em relação à P, para todo $\alpha>0$, $P \in \Delta_n$.

Prova:

Concavidade implica em pseudoconcavidade. Assim temos que provar o teorema somente para o caso $\alpha>1$, pois para $0<\alpha\leq 1$, $H_\alpha(P)$ é côncava com respeito à P, para todo $P \in \Delta_n$, $n \geq 2$.

$$\nabla H_\alpha(P)(Q-P) = \frac{\alpha}{(1-\alpha)\ell n2} \frac{\sum_{i=1}^{n} p_i^{\alpha-1}}{\sum_{i=1}^{n} p_i^{\alpha}} (q_i - p_i)$$

$$= \frac{\alpha}{(1-\alpha)\ell n2 \sum_{i=1}^{n} p_i^{\alpha}} \left(\sum_{i=1}^{n} p_i^{\alpha-1} q_i - \sum_{i=1}^{n} p_i^{\alpha} \right) \tag{2.30}$$

De modo que $\nabla H_\alpha(P)(Q-P) \leq 0$, se e somente se,

$$\sum_{i=1}^{n} p_i^{\alpha-1} q_i \geq \sum_{i=1}^{n} p_i^{\alpha} \tag{2.31}$$

Por outro lado $H_\alpha(Q) \leq H_\alpha(P)$, para $\alpha>1$, se e somente se,

$$\sum_{i=1}^{n} q_i^{\alpha} \geq \sum_{i=1}^{n} p_i^{\alpha} \;. \tag{2.32}$$

O teorema segue imediatamente do seguinte lema:

<u>LEMA AUXILIAR</u>: Para todo $\alpha > 0$, $P,Q \in \Delta_n$,

$$\sum_{i=1}^{n} q_i^{\alpha}\, p_i^{\alpha-1} \geq \sum_{i=1}^{n} p_i^{\alpha} \quad \text{implica}$$

$$\sum_{i=1}^{n} q_i^{\alpha} \geq \sum_{i=1}^{n} p_i^{\alpha}$$

Prova:

Para cada par (p_i, q_i) existe um $\varepsilon_i > -1$ tal que $q_i = p_i(1 + \varepsilon_i)$ e $\sum_{i=1}^{n} p_i\, \varepsilon_i = 0$

De fato:

$$1 = \sum_{i=i}^{n} q_i = \sum_{i=1}^{n} p_i(1+\varepsilon_i) = \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} p_i \varepsilon_i = 1 + \sum_{i=1}^{n} p_i \varepsilon_i$$

donde $\sum_{i=1}^{n} p_i\, \varepsilon_i = 0$

Agora se $\sum_{i=1}^{n} q_i\, p_i^{\alpha-1} \geq \sum_{i=1}^{n} p_i^{\alpha}$, isto é,

$$\sum_{i=1}^{n} q_i\, p_i^{\alpha-1} = \sum_{i=1}^{n} p_i(1+\varepsilon_i)p_i^{\alpha-1} = \sum_{i=1}^{n} p_i^{\alpha} + \sum_{i=1}^{n} p_i^{\alpha}\, \varepsilon_i \geq \sum_{i=1}^{n} p_i^{\alpha}$$

então $\sum_{i=1}^{n} p_i^{\alpha}\, \varepsilon_i \geq 0$. Usando a expansão da série de Taylor de $(1+\varepsilon_i)^{\alpha}$ para $\alpha > 1$, obtemos:

$$(1 + \varepsilon_i)^\alpha > 1 + \alpha\varepsilon_i, \quad \varepsilon_i > -1, \quad \alpha > 1$$

de modo que

$$\sum_{i=1}^{n} q_i^\alpha = \sum_{i=1}^{n} p_i^\alpha (1+\varepsilon_i)^\alpha \geqslant \sum_{i=1}^{n} (1+\alpha\varepsilon_i) p_i^\alpha = \sum_{i=1}^{n} p_i^\alpha + \alpha \sum_{i=1}^{n} \varepsilon_i p_i^\alpha$$

$$\geqslant \sum_{i=1}^{n} p_i^\alpha \quad .$$

OBSERVAÇÃO:

Da definição de pseudoconcavidade segue que uma condição necessária e suficiente para um ponto $\bar{x}$ ser o ponto máximo de θ sobre um conjunto convexo Γ é que $\nabla\theta(\bar{x}) = 0$. É bem conhecido que funções côncavas possuem esta propriedade, mas existem funções pseudocôncavas que não são côncavas.

Usando teoremas examinados em [16] (capítulo 10) para otimização forçada de funções pseudocôncavas concluímos que $p^\circ = (\frac{1}{n},\ldots,\frac{1}{n})$ é o único ponto de máximo de $H_\alpha(P)$ sobre Δ_n, para todo $\alpha > 0$, $n \geqslant 2$.

§ 2.2 - Entropia de Ordem α e a Probabilidade de Erro

A - Cota Superior H_α

Em [13], [10] e [14] uma cota superior sobre a probabilidade de erro foi obtida em termos da entropia de Shannon.

$$P(e) \leqslant \frac{1}{2} \; Ex \left[H(C/x)\right] \qquad (2.33)$$

onde Ex $[H(C/x)]$ é a equivocação e $H(C/x)$ é a entropia condicional de Shannon de C dado que X=x e é dada por:

$$H(C/x) = - \sum_{i=1}^{n} P(C_i/x) \log P(C_i/x)$$

Esta cota é baseada na seguinte desigualdade:

$$1 - p_M \leqslant - \frac{1}{2} \sum_{i=1}^{n} p_i \log p_i = \frac{1}{2} H(P); \quad P \in \Delta_n \qquad (2.34)$$

No que segue, (2.34) é extendida para o caso $\alpha > 1$. Então usando o fato que $H_\alpha(P)$ é uma função decrescente de α, uma cota mais próxima da probabilidade de erro é obtida.

TEOREMA 2.3 - Para todo $n \geqslant 2$ e $P \in \Delta_n$

a) $1-p_M \leqslant \frac{1}{2} H_\alpha(P)$, para todo $\alpha > 0$

se $\frac{1}{n} < p_M \leqslant \frac{1}{2}$

b) Existe um $\alpha(P)$ tal que:

$1-p_M > \frac{1}{2} H_\alpha(P)$ para todo $\alpha > \alpha(P)$

se $\frac{1}{2} < p_M \leqslant 1$.

Prova:

Pelo lema 2.2, $\lim\limits_{\alpha\to\infty} H_\alpha(P) = H_\infty(P) = -\log p_M$.

Consideremos a função:

$$f(p_M)=1-p_M + \frac{1}{2}\log p_M = 1-p_M + \frac{1}{2\ell n 2}\ell n\, p_M$$

com

$$f'(p_M) = -1 + \frac{1}{2\ell n 2}\cdot\frac{1}{p_M}$$

e

$$f''(p_M) = \frac{-1}{2\ell n 2}\cdot\frac{1}{p_M^2} .$$

Desde que $f''(p_M) < 0$, para todo p_M, $f(p_M)$ é uma função estritamente côncava e atinge seu único máximo em $f'(p_M)=0$, isto é,

$$\frac{1}{2\ell n 2}\cdot\frac{1}{p_M} = 1 \Longrightarrow p_M = \frac{1}{2\ell n 2} \simeq 0,72 .$$

Além disso:

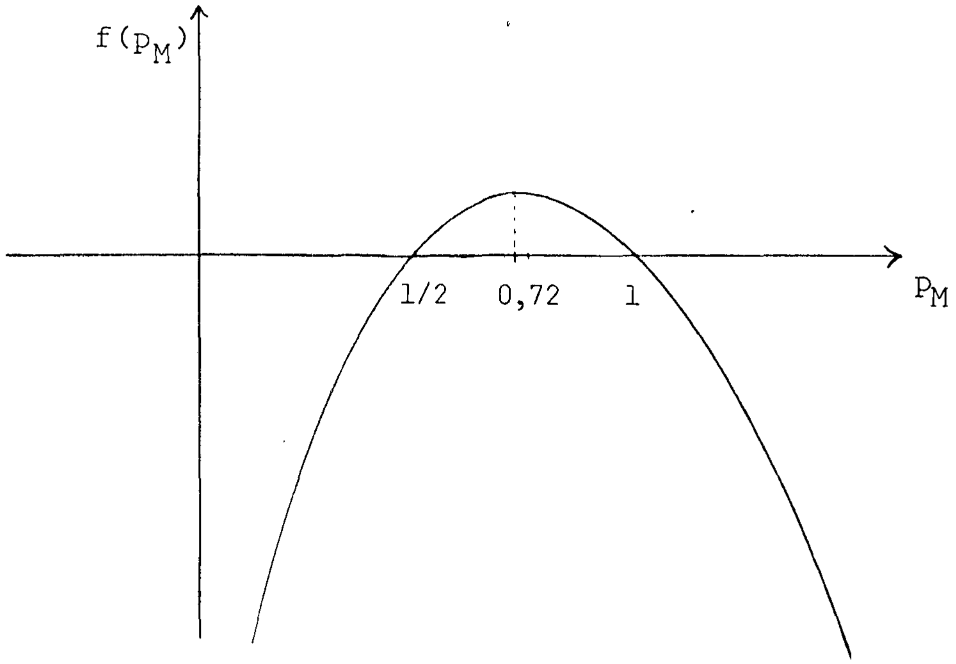
$$f(\tfrac{1}{2}) = 1 - \frac{1}{2} + \frac{1}{2}\log\frac{1}{2} = 1 - \frac{1}{2} - \frac{1}{2} = 0$$

$$f(1)=1-1 + \frac{1}{2}\log 1 = 1-1+0 = 0,$$

estes são os únicos zeros de $f(p_M)$ pois $f(p_M)$ é côncava.

$$\text{Como} \lim_{p_M \to 0} f(p_M) = \lim_{p_M \to 0} \left[1-p_M + \frac{1}{2} \log p_M\right] = -\infty$$

o gráfico de $f(p_M)$ é como mostra a figura abaixo:



Portanto:

$$1-p_M \leq \frac{1}{2} H_\infty(P), \quad \text{para} \quad \frac{1}{m} < p_M \leq \frac{1}{2} \tag{2.35}$$

$$1-p_M > \frac{1}{2} H_\infty(P), \quad \text{para} \quad \frac{1}{2} < p_M \leq 1 \tag{2.36}$$

De fato:

$$\text{Para } \frac{1}{m} < p_M \leq \frac{1}{2}, \quad f(p_M) \leq 0 \Longrightarrow$$

$$\Longrightarrow 1-p_M + \frac{1}{2} \log p_M \leq 0 \Longrightarrow 1-p_M \leq -\frac{1}{2} \log p_M \Longrightarrow$$

$$\Longrightarrow 1-p_M \leq \frac{1}{2} \lim_{\alpha \to \infty} H_\alpha(P) = \frac{1}{2} H_\infty(P).$$

Para $\frac{1}{2} < p_M \leqslant 1, \quad f(p_M) \geqslant 0 \Longrightarrow$

$$\Longrightarrow 1-p_M + \frac{1}{2} \log p_M \geqslant 0 \Longrightarrow 1-p_M \geqslant -\frac{1}{2} \log p_M \Longrightarrow$$

$$\Longrightarrow 1-p_M \geqslant \frac{1}{2} \lim_{\alpha\to\infty} H_\alpha(P) = \frac{1}{2} H_\infty(P).$$

Desde que $H_\alpha(P)$ é uma função decrescente de α, deduzimos (a) de (2.35), isto é,

se $\frac{1}{m} < p_M \leqslant \frac{1}{2}$, $1-p_M \leqslant \frac{1}{2} H_\infty(P) \leqslant \frac{1}{2} H_\alpha(P)$ para todo $\alpha > 0$.

Desde que $H_\alpha(P)$ é uma função estritamente decrescente de α, para $P \neq P^o$, (2.36) será válida para todo $\alpha > \alpha^*(P)$ onde

$$\alpha^*(P) = \inf\{\alpha / 1-p_M > \frac{1}{2} H_\alpha(P)\}.$$

assim provando (b).

De fato: Se $\frac{1}{2} < p_M \leqslant 1 \Longrightarrow 1-p_M > \frac{1}{2} H_\infty(P) =$

$$= \frac{1}{2} H_{\alpha^*(P)}(P) > \frac{1}{2} H_\alpha(P) \text{ para todo } \alpha > \alpha^*(P).$$

<u>OBSERVAÇÃO</u>: A quantidade $\alpha^*(P)$ depende também de "n" mas não indicaremos isto afim de simplificar a notação.

Os seguintes teorema e lema provam que $\alpha^*=2$ onde

$$\alpha^* = \inf \{\alpha^*(P)/P \in \Delta_n, n \geqslant 2\}$$

TEOREMA 2.4 - Para todo $P \in \Delta_n$, $n \geq 2$ e $0 \leq \alpha \leq 2$.

$$1-p_M \leq \frac{1}{2} H_\alpha(P) \tag{2.37}$$

Prova:

O teorema 2.3 afirma que (2.37) é válida para todo $P \in \Delta_n$, $n \geq 2$ tal que $p_M \leq \frac{1}{2}$. Assim nos limitaremos ao caso quando $p_M > \frac{1}{2}$. É suficiente provar o teorema para $\alpha = 2$. Provaremos (2.37) para $n = 2$ e extenderemos então a $n > 2$.

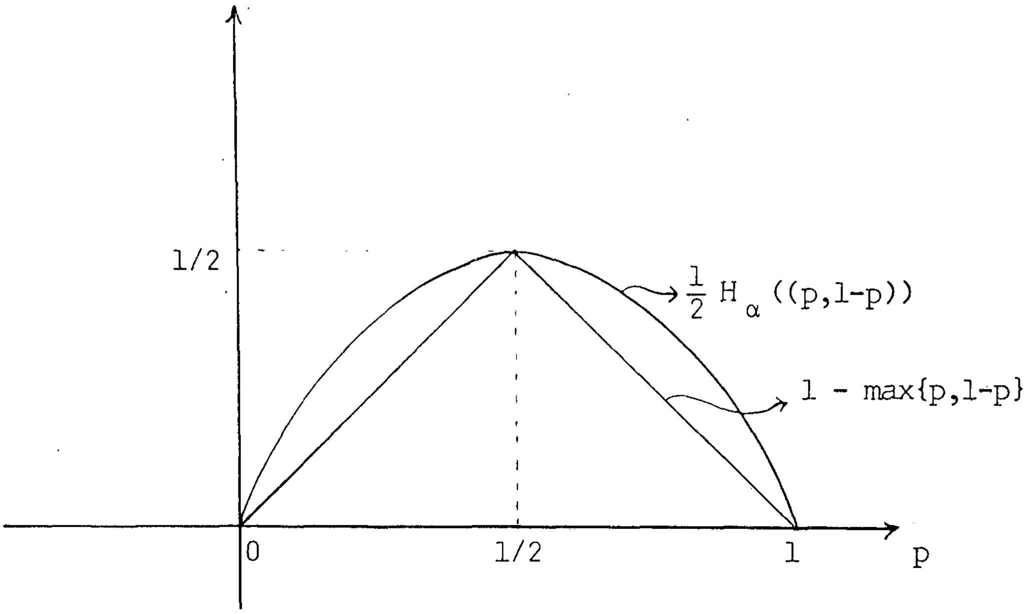
Para o caso $n = 2$, de (2.4) e (2.3) temos:

$$\frac{1}{2} H_\alpha(1,0) = 0$$

$$\frac{1}{2} H_\alpha(1/2,\ 1/2) = \frac{1}{2}\left(-\log \frac{1}{2}\right) = \frac{1}{2}$$

$$\frac{1}{2} H_\alpha(0,1) = 0$$

Desde que o gráfico de $1-\max\{p,\ 1-p\}$, $0 \leq p \leq 1$ consiste de duas retas entre $(0,0)$ e $(1/2,\ 1/2)$ e entre $(1/2,1/2)$ e $(1,0)$ obtemos o resultado desejado da concavidade de $H_\alpha(p,1-p)$ para $\alpha \leq 2$, $0 \leq p \leq 1$.



Seja $P \in \Delta_n$, $n > 2$, tal que $p_M > \frac{1}{2}$ e, sem perda de generalidade, assumimos que $p_M = p_n$.

De (2.6) temos que $\quad H_\alpha((1-p_M,p_M)) \leqslant H_\alpha(P) \qquad (2.38)$

e desde que o teorema é válido para $n = 2$, temos:

$$1-p_M \leqslant \frac{1}{2} H_\alpha((1-p_M,p_M)) \tag{2.39}$$

De (2.38) e (2.39) obtemos o resultado desejado, ou seja,

$$1-p_M \leqslant \frac{1}{2} H_\alpha(P) \quad \text{para todo} \quad P \in \Delta_n,\ n \geqslant 2 \quad \text{e} \quad 0 < \alpha \leqslant 2.$$

<u>LEMA 2.4</u> - Para todo $\alpha > 2$ e $n \geqslant 2$ existe um conjunto não vazio $K \subset \Delta_n$

tal que:

$$1-p_M > \frac{1}{2} H_\alpha(P), \text{ para todo } P \in K \tag{2.40}$$

<u>OBSERVAÇÃO</u>: Pelo teorema 2.3, todo $P \varepsilon K$ é tal que $p_M > \frac{1}{2}$.

Prova:

É suficiente provar o teorema para n=2 desde que o caso $n > 2$ pode ser obtido do caso n=2 adicionando zeros.

Dado $\alpha > 2$, denotemos p', p" pontos onde $\frac{1}{2} H_\alpha((p,1-p))$, muda de convexa para côncava e de côncava para convexa. As relações (2.26) e (2.27) implicam que $0 < p' < \frac{1}{2}$ e $\frac{1}{2} < p'' < 1$.

Portanto para todo $P \varepsilon K$

$$1 - p_M > \frac{1}{2} H_\alpha(P) \tag{2.41}$$

onde

$$K=\{(p,1-p) : 0 < p < p' \quad \text{ou} \quad p'' < p < 1\}$$

<u>OBSERVAÇÃO</u>: A quantidade p' não é necessariamente o limite superior para o conjunto de pontos onde (2.41) é válida.

Notemos que o teorema 2.3 vale em particular para C/x, para todo x, de modo que:

$$1\text{-}\max\{P(C_1/x),\ldots,P(C_n/x)\} \leqslant \frac{1}{2} H_2(C/x) \tag{2.42}$$

onde

$H_\alpha(C/x)=\frac{1}{1-\alpha} \log\left[\sum_{i=1}^{n} P^2(C_i/x)\right]$, $\alpha > 0$, é a entropia condicional de ordem α de C dado que X=x.

Tomando expectativas com respeito a x de ambos os lados de (2.42) obtemos:

$$P_e \leqslant \frac{1}{2} I_2 \qquad (2.43)$$

onde,

$$P_e = E_x\left[1-\max\{P(C_1/x),\ldots,P(C_n/x)\}\right]$$

$$I_\alpha = E_x\left[H_\alpha(C/x)\right], \qquad \alpha > 0$$

A cota (2.43) é mais aproximada da igualdade do que (2.33) desde que H_α é uma função decrescente de α. Se supormos que para todos os valores de x, $P(e/x) \leqslant \frac{1}{2}$ obtemos uma cota mais geral igual à

$$P_e \leqslant \frac{1}{2} I_\alpha \qquad \text{para todo} \quad \alpha > 0, \qquad (2.44)$$

para o caso em que n=2, isto é, para uma distribuição de duas probabilidades $P(e/x) \leqslant \frac{1}{2}$, para todo x, portanto a última cota nunca é válida para $\alpha > 2$. O lado direito de (2.43) será chamado cota I_2, enquanto o lado direito de (2.44) será chamado cota I_α.

Segue o lema 2.4 que a extensão de (2.37) para $\alpha > 2$ e todo $P \in \Delta_n$ é inválida, contudo, uma cota similar à de (2.37) para todo $\alpha > 2$, $n \geqslant 2$, e $P \in \Delta_n$ existe mas com um fator numérico de $\ell n2 \cong 0{,}693$ em vez de 0,5.

<u>TEOREMA 2.5</u> - Para todo $n \geq 2$ e $\alpha > 0$,

$$1-p_M \leq (\ell n2)H_\alpha(P), \text{ para todo } P \in \Delta_n \qquad (2.45)$$

Prova:

A demonstração é baseada na seguinte desigualdade

$$\ell n x \leq x - 1 \quad \text{para todo} \quad x > 0$$

De (2.5) temos que $\lim\limits_{\alpha\to\infty} H_\alpha(P) = -\log p_M = \dfrac{-\ell n p_M}{\ell n 2}$.

Invertendo a desigualdade (2.46) obtemos

$$1-p_M \leq -\ell n p_M = -(\ell n2) \log p_M = (\ell n2)H_\infty(P). \qquad (2.47)$$

Desde que $H_\alpha(P)$ é uma função decrescente de α (2.47) implica em (2.45).

B - <u>Cotas Inferiores H_α</u>

Nesta seção cotas inferiores sobre a probabidade de erro derivadas da entropia de ordem α serão desenvolvidas por analogia à cota de Fano.

Definimos

$$\bar{h}_\alpha(p_e) = \sup_P \{H_\alpha(P)/P \in \Delta_n,\ 1-\max\{p_1,\dots,p_n\} = p_e\} \qquad (2.48)$$

<u>TEOREMA 2.6</u> - Para $\alpha \neq 1$, o máximo $\bar{h}_\alpha(p_e)$ é atingido quando todos p_i, $i=1,2,\dots,n$ (exceto talvez um) são iguais a $p_e/(n-1)$ e então

$$\bar{h}_\alpha(p_e) = \frac{1}{1-\alpha} \log\left[(1-p_e)^\alpha + \left(\frac{1}{n-1}\right)p_e^\alpha\right] \tag{2.49}$$

Prova:

Seja $P \in \Delta_n$ com $p_e = 1-\max\{p_1, p_2, \dots, p_n\}$ e, sem perda de generalidade, suponhamos que $\max\{p_1,\dots,p_n\} = p_n$. Então

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\left[\sum_{i=1}^{n-1} p_i^\alpha + p_n^\alpha\right] =$$

$$= \frac{1}{1-\alpha} \log\left[(n-1) \sum_{i=1}^{n-1} \frac{1}{(n-1)} p_i^\alpha + p_n^\alpha\right]$$

<u>Caso 1</u>: Quando $\alpha > 1$

Desde que a função $w(x)=x$, $x > 0$ é estritamente convexa para $\alpha > 1$ e $\sum_{i=1}^{n-1} \frac{1}{(n-1)} = 1$, temos:

$$\sum_{i=1}^{n-1} \frac{1}{(n-1)} p_i^\alpha \geq \left[\sum_{i=1}^{n-1} \frac{1}{(n-1)} p_i\right]^\alpha = \left[\frac{p_e}{n-1}\right]^\alpha \tag{2.50}$$

com igualdade se e somente se $p_i = \frac{p_e}{n-1}$, para todo $i=1,2,\dots,n-1$. Desde que log é uma função estritamente crescente e $\frac{1}{1-\alpha} < 0$ para $\alpha > 1$ concluimos que:

$$H_\alpha(P) \leqslant \frac{1}{1-\alpha} \log\left[(1-p_e)^\alpha + (n-1)\left[\frac{p_e}{n-1}\right]^\alpha\right]$$

$$= \frac{1}{1-\alpha} \log\left[(1-p_e)^\alpha + \left[\frac{1}{n-1}\right]^{\alpha-1} p_e^\alpha\right] \tag{2.51}$$

De fato:

$$(2.50) \left[\frac{p_e}{n-1}\right]^\alpha \leqslant \sum_{i=1}^{n-1} p_i^\alpha \text{ , isto é,}$$

$$(n-1)\left[\frac{p_e}{n-1}\right]^\alpha \leqslant (n-1) \sum_{i=1}^{n-1} p_i^\alpha$$

logo: $p_n^\alpha + (n-1)\left[\frac{p_e}{n-1}\right]^\alpha \leqslant (n-1) \sum_{i=1}^{n-1} p_i^\alpha + p_n^\alpha = \sum_{i=1}^{n} p_i^\alpha$

Desde que $p_n = \max\{p_1,\ldots,p_n\}$ e $p_e = 1-\max_i p_i$ temos que $p_n = 1-p_e$ logo:

$$(1-p_e)^\alpha + (n-1)\left[\frac{p_e}{n-1}\right]^\alpha \leqslant \sum_{i=1}^{n} p_i^\alpha \text{ , ou seja,} \tag{2.52}$$

$$\log\left[(1-p_e)^\alpha + (n-1)\left[\frac{p_e}{n-1}\right]^\alpha\right] \leqslant \log \sum_{i=1}^{n} p_i^\alpha \text{ , isto é}$$

$$\frac{1}{1-\alpha} \log\left[(1-p_e)^\alpha + (n-1)\left[\frac{p_e}{n-1}\right]^\alpha\right] > \frac{1}{1-\alpha} \log \sum_{i=1}^{n} p_i^\alpha$$

portanto: $H_\alpha(P) \leqslant \frac{1}{1-\alpha} \log\left[(1-p_e)^\alpha + (\frac{1}{1-\alpha})^{\alpha-1} p_e^\alpha\right]$

Caso 2: $0 < \alpha < 1$

Para este caso $w(x)=x^{\alpha}$, $x > 0$, é estritamente côncava, e portanto (2.50) e (2.52) ficam invertidas, mas $\frac{1}{1-\alpha} > 0$, para $0 < \alpha < 1$, assim obtemos novamente (2.51). Isto completa a demonstração.

Note que quando $\alpha \to 1$ temos:

$$H_1(P) \leq -(1-p_e)\log(1-p_e)-p_e \log \frac{p_e}{n-1}, \tag{2.53}$$

que a bem conhecida cota de FANO.

Casos Particulares

Notemos que (2.52) é válida para $\alpha > 1$, enquanto que para $\alpha < 1$ a desigualdade (2.52) é invertida. A solução explícita de (2.52) em termos de p_e não é possível para α arbitrário, contudo para $\alpha=2$ obtemos:

$$\frac{n-1}{n}\left[1 - \sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n - 1}}\right] \leqslant p_e. \tag{2.54}$$

De fato: Fazendo $\alpha=2$ em (2.52) obtemos:

$$\sum_{i=1}^{n} p_i^2 \geqslant (1 - p_e)^2 + \left(\frac{1}{n-1}\right) p_e^2 \tag{2.55}$$

$$= 1 - 2 p_e + \left(1 + \frac{1}{n-1}\right)p_e^2$$

$$= 1 - 2\, p_e + \frac{n}{n-1}\, p_e^2 \quad ,$$

portanto:

$$n \sum_{i=1}^{n} p_i^2 \geqslant n - 2n\, p_e + \frac{n^2}{n-1}\, p_e^2$$

$$n \sum_{i=1}^{n} p_i^2 - 1 \geqslant (n-1) - 2n\, p_e + \frac{n^2}{n-1}\, p_e^2$$

$$\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n-1} \geqslant 1 - 2\, \frac{n}{n-1}\, p_e + \frac{n^2}{n-1}\, p_e^2$$

$$\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n-1} \geqslant \left(1 - \frac{n}{n-1}\, p_e\right)^2$$

$$\sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n-1}} \geqslant 1 - \frac{n}{n-1}\, p_e$$

$$-1 + \sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n-1}} \geqslant - \frac{n}{n-1}\, p_e$$

$$1 - \sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n - 1}} \leqslant \frac{n}{n-1} P_e$$

$$\frac{n-1}{n}\left[1 - \sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n - 1}}\right] \leqslant P_e$$

Fazendo $p_i = P(C_i/x)$ e tomando o valor esperado de ambos os lados de (2.54), obtemos uma cota inferior para P_e, a saber:

$$E_x\left\{\frac{n-1}{n}\left[1 - \sqrt{\frac{n \sum_{i=1}^{n} P^2(C_i/x) - 1}{n - 1}}\right]\right\} \leqslant P_e \qquad (2.56)$$

As cotas (2.54) e (2.56) estão implícitas numa vizinhança dada por Cover e Hart [5]. Também coincide com o resultado de Devijver [7] para a distância Bayesiana. Para o caso de duas probabilidades, isto é, quando n=2, Toussaint [27] obteve um resultado similar à (2.56). Aplicando a desigualdade de Jensen [7] para a função côncava $\sqrt{}$ obtemos:

$$\frac{n-1}{n}\left\{1 - \sqrt{\frac{n\, E_x\left[\sum_{i=1}^{n} P^2(C_i/x) - 1\right]}{n - 1}}\right\} \leqslant P_e \qquad (2.57)$$

Claramente (2.56) é mais próxima da igualdade que (2.57).

C A P Í T U L O   I I I

ENTROPIA PARAMÉTRICA GENERALIZADA E A PROBABILIDADE DE ERRO

Neste capítulo estudaremos algumas novas propriedades da entropia generalizada paramétrica chamada entropia de de ordem α e grau β. Foram obtidas cotas superiores e inferiores sobre a probabilidade de erro em termos da entropia paramétrica, resultados estes que constituem basicamente o desenvolvimento original deste trabalho.

§ 3.1 - Entropia Generalizada Paramétrica

A entropia generalizada paramétrica ou entropia de ordem α e grau β para uma distribuição completa de n-probabilidades é dada por:

$$H_\alpha^\beta(P)=H_\alpha^\beta(p_1,\ldots,p_n)=(2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}}-1\right], \qquad (3.1)$$

$$\alpha,\beta > 0;\ \alpha \neq 1,\ \beta \neq 1,$$

$P=(p_1,\ldots,p_n) \in \Delta_n$, onde Δ_n foi definido em (1.1).

O seguinte teorema foi demonstrado em [11] e será usado na demonstração do lema a seguir.

rad

TEOREMA 3.1 - Se $a=(a_1,\dots,a_n)$ é uma série de números reais positivos e $p=(p_1,\dots,p_n)$ uma série de pesos a média $M_n^r(a,p)$ de ordem r é definida por:

$$M_n^r(a,p) = \left[\frac{\sum_{i=1}^{n} p_i a_i^r}{\sum_{i=1}^{n} p_i} \right]^{1/r} \quad ; \; r\neq 0, \; |r| < +\infty \tag{3.2}$$

é uma função estritamente crescente de r quando os a_i não são todos iguais, caso contrário a média é o valor comum.

a) Quando r=0 temos:

$$M_n^o(a,p) = \left[\prod_i a_i^{p_i} \right]^{\frac{1}{\sum_{i=1}^{n} p_i}} \tag{3.3}$$

b) Quando $r \to -\infty$ temos:

$$M_n^{-\infty}(a,p) = \min_i \{a_1,a_2,\dots,a_n\} \tag{3.4}$$

c) Quando $r \to +\infty$ temos:

$$M_n^{+\infty}(a,p) = \max_i \{a_1,a_2,\dots,a_n\} \tag{3.5}$$

LEMA 3.1 - (Thierry Van der Pyl [29]): Para todo $P \in \Delta_n^o$ e $\beta > 0$ fixo, $H_\alpha^\beta(P)$ é uma função decrescente de α.

Prova:

Seja $P=(p_1,\ldots,p_n) \in \Delta_n^o$, onde Δ_n^o é definido em (1.2).

Em (3.2), façamos $a=p_i^{\beta-1}$, $p=P$ e $r=\frac{\alpha-1}{\beta-1}$, assim:

$$M_n^r(a,p)=\left[\sum_{i=1}^{n} p_i (p_i^{\beta-1})^{\frac{\alpha-1}{\beta-1}}\right]^{\frac{\beta-1}{\alpha-1}} = \left[\sum_{i=1}^{n} p_i^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}}$$

Suponhamos β fixo, isto é, os $(a_i)_i$ são fixos, de acordo com o teorema (3.1), $M_n^r(a,p)$ é uma função crescente de r, portanto uma função crescente de α quando $\beta > 1$, uma função decrescente de α quando $\beta < 1$.

Consideremos os dois casos seguintes:

<u>1º Caso: $\beta > 1$</u>

Desde que $M^r(a,p)$ é uma função crescente de α quando $\beta > 1$, para $\alpha_1 \geqslant \alpha_2$ temos:

$$\left[\sum_{i=1}^{n} p_i^{\alpha_1}\right]^{\frac{\beta-1}{\alpha_1-1}} \geqslant \left[\sum_{i=1}^{n} p_i^{\alpha_2}\right]^{\frac{\beta-1}{\alpha_2-1}} \tag{3.6}$$

Subtraindo 1 (um) à ambos os membros da desigualdade e multiplicando a desigualdade resultante por $\left[(2^{1-\beta}-1)^{-1} < 0,\ \beta > 1\right]$, obtemos:

$$(2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{\alpha_1}\right)^{\frac{\beta-1}{\alpha_1-1}}-1\right] \leq (2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^{\alpha_2}\right)^{\frac{\beta-1}{\alpha_2-1}}-1\right] \quad (3.7)$$

isto é: $H^{\beta}_{\alpha_1}(P) \leq H^{\beta}_{\alpha_2}(P)$. (3.8)

2º Caso: $\beta < 1$

Para este caso $M^r(a,p)$ é uma função decrescente de α, portanto para $\alpha_1 \geq \alpha_2$ (3.6) fica invertida, assim, sendo $(2^{1-\beta}-1) > 0$, para $\beta < 1$ obtemos novamente (3.7) e consequentemente (3.8); completando a demonstração do lema proposto.

LEMA 3.2 - a) $H^{\beta}_{0}(P)=(2^{1-\beta}-1)^{-1}\left[n^{1-\beta}-1\right]$, para todo $P\varepsilon\Delta_n$ e $\beta>0$ (3.9)

b) $H^{\beta}_{\alpha}(P^o)=(2^{1-\beta}-1)^{-1}\left[n^{1-\beta}-1\right]$, para todo $\alpha,\beta>0$, (3.10)

onde P^o é definido em (1.3).

c) $H^{\beta}_{\alpha}(P^i)=0$, para todo $\alpha,\beta>0$; $i=1,2,\ldots,n$; (3.11)

onde P^i é definido em (1.4).

d) $\lim\limits_{\alpha\to 0} H^{\beta}_{\alpha}(P)=H^{\beta}_{\infty}(P)=(2^{1-\beta}-1)^{-1}\left[(p_M)^{\beta-1}-1\right]$ (3.12)

para todo $P \varepsilon \Delta_n$, $\beta > 0$

e) $H_\alpha^\beta(1-p_M,p_M) \leqslant H_\alpha^\beta(P)$, para todo $P \varepsilon \Delta_n$, $\beta > 0$ (3.13)

onde $p_M = \max\limits_i p_i$

Prova:

a) $$H_0^\beta(P)=(2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^0\right)^{\frac{\beta-1}{0-1}}-1\right] = (2^{1-\beta}-1)^{-1}\left[n^{1-\beta}-1\right]$$

b) $$H_\alpha^\beta(P^o)=H_\alpha^\beta((\frac{1}{n},\ldots,\frac{1}{n}))=(2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n}(\frac{1}{n})^\alpha\right)^{\frac{\beta-1}{\alpha-1}}-1\right]$$

$$= (2^{1-\beta}-1)^{-1}\left[(n.n^{-\alpha})^{\frac{\beta-1}{\alpha-1}}-1\right] = (2^{1-\beta}-1)^{-1}\left[n^{1-\beta}-1\right]$$

c) $$H_\alpha^\beta(P^i)=H_\alpha^\beta((0,\ldots,0,1,0,\ldots,0))=(2^{1-\beta}-1)^{-1}\left[(1)^{\frac{\beta-1}{\alpha-1}}-1\right]=(2^{1-\beta}-1)^{-1}.0=0$$

d) A informação de ordem α e do tipo β pode também ser escrita da seguinte maneira:

$H_\alpha^\beta(P) = g_\beta(H_\alpha(P))$; onde $P \varepsilon \Delta_n$

$H_\alpha^\beta(P)$ é a informação de ordem α e:

$$g_\beta(x) = (2^{1-\beta}-1)^{-1}\left[2^{(1-\beta)x}-1\right]$$

De fato:

$$g_\beta(H_\alpha(P))=(2^{1-\beta}-1)\left[2^{(1-\beta)\frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n} p_i^\alpha\right)}-1\right]$$

$$= (2^{1-\beta}-1)^{-1}\left\{\left[2^{\log\left(\sum_{i=1}^{n} p_i^\alpha\right)}\right]^{\frac{\beta-1}{\alpha-1}}-1\right\}$$

$$= (2^{1-\beta}-1)^{-1}\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}}-1\right] = H_\alpha^\beta(P).$$

Como já vimos anteriormente em (2.5)

$$\lim_{\alpha\to\infty} H\ (P) = -\log p_M \quad \text{para todo } P \in \Delta_n.$$

Desde que g é contínua para $\beta \neq 1$ temos:

$$\lim_{\alpha\to\infty} H_\alpha^\beta(P) = \lim_{\alpha\to\infty} g_\beta(H_\alpha(P))$$

$$= g_\beta(\lim_{\alpha\to\infty} H_\alpha(P)) = g_\beta(-\log p_M) =$$

$$= (2^{1-\beta}-1)^{-1}\left[2^{(1-\beta)(-\log p_M)}-1\right] =$$

$$= (2^{1-\beta}-1)^{-1}\left[(2^{\log p_M})^{\beta-1}-1\right] =$$

$$= (2^{1-\beta}-1)^{-1}\,(p_M^{\beta-1}-1).$$

e) Do lema 2.3 temos que: Para todo $P \in \Delta_n$, $p_M = \max_i p_i$

$$H_\alpha(1-p_M,\ p_M) \leqslant H_\alpha(P), \tag{3.14}$$

Desde que a função g definida anteriormente é uma função crescente de x, obtemos de (3.14):

$$g_\beta\left[H_\alpha(1-p_M,p_M)\right] \leqslant g_\beta\left[H_\alpha(P)\right]$$

portanto:

$$H_\alpha^\beta(1-p_M,p_M) \leqslant H_\alpha^\beta(P), \text{ para todo } P \in \Delta_n$$

$$p_M = \max_i p_i$$

Assim completamos a demonstração do lema.

<u>TEOREMA 3.2 [29]</u>: a) Para $P \in \Delta_n$ e $0 < \alpha \leqslant \frac{1}{2-\beta}$, $H_\alpha^\beta(P)$ é uma função côncava com respeito à P.

b) Para todo $P \in \Delta_n$, $\alpha > 0$ e $\beta = 2$, $H_\alpha^\beta(P)$ é uma função côncava com respeito à P.

Prova:

Lembre a desigualdade de Minkowski [11]:

Sejam $a_i \geqslant 0$, $b_i \geqslant 0$, $i=1,\dots,n$ números reais.

. Se $r > 1$,

$$\left(\sum_{i=1}^{n} a_i^r\right)^{1/r} + \left(\sum_{i=1}^{n} b_i^r\right)^{1/r} \geq \left(\sum_{i=1}^{n} (a_i+b_i)^r\right)^{1/r} \tag{3.15}$$

e

$$\left(\sum_{i=1}^{n} a_i^{1/r}\right)^{r} + \left(\sum_{i=1}^{n} b_i^{1/r}\right)^{r} \leq \left(\sum_{i=1}^{n} (a_i+b_i)^{1/r}\right)^{r} \tag{3.16}$$

. Se $0 < r < 1$, as desigualdades (3.15) e (3.16) ficam invertidas.

Sejam $P=(p_1,\ldots,p_n) \in \Delta_n$

$Q=(q_1,\ldots,q_n) \in \Delta_n$

$\lambda,\mu \in [0,1]$, tais que $\lambda + \mu = 1$

e $\lambda P + \mu Q = (\lambda p_1+\mu q_1,\ldots,\lambda p_n+\mu q_n) \in \Delta_n$

Aplicando as relações (3.15) e (3.16) ao caso:

$$a_i = \lambda p_i \geq 0,$$

$$b_i = \mu q_i \geq 0, \quad \text{e} \quad r = \alpha.$$

Na demonstração do item (a) distinguiremos três casos:

1º Caso: $\alpha > 1$

Temos a seguinte desigualdade:

$$\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha} \geqslant \left[\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{\alpha}\right]^{1/\alpha} .$$

. Se $\alpha\,\dfrac{\beta-1}{\alpha-1} > 0$, isto é, $\beta > 1$,

$$\left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} \geqslant \left[\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} . \qquad (3.17)$$

Se além disso, $\alpha\,\dfrac{\beta-1}{\alpha-1} > 1$, isto é, $\beta > 2 - \dfrac{1}{\alpha}$, a função $Y = x^{\alpha\frac{\beta-1}{\alpha-1}}$ é convexa, donde:

$$\lambda\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} + \mu\left[\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} \geqslant \left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} . \qquad (3.18)$$

pois f é convexa se para $0 \leqslant \lambda \leqslant 1$, tivermos:

$f(\lambda x+(1-\lambda)y) \leqslant \lambda f(x)+(1-\lambda)f(y)$. Caso esta desigualdade seja invertida f será côncava.

De (3.17) e (3.18) obtemos então:

$$\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} > \left[\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} \tag{3.19}$$

Subtraindo $1=\lambda+\mu$ à ambos os membros de (3.19) e dividindo a desigualdade resultante por $2^{1-\beta}-1$, obtemos a nova desigualdade:

$$\frac{\lambda}{2^{1-\beta}-1}\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} -1\right] + \frac{\mu}{2^{1-\beta}-1}\left[\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} -1\right] \leqslant$$

$$\leqslant \frac{1}{2^{1-\beta}-1}\left[\left(\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{\alpha}\right)^{\frac{\beta-1}{\alpha-1}} -1\right] \tag{3.20}$$

ou seja:

$$\lambda H_{\alpha}^{\beta}(P) + \mu H_{\alpha}^{\beta}(Q) \leqslant H_{\alpha}^{\beta}(\lambda P+\mu Q) \tag{3.21}$$

para $\alpha > 1$, $\beta > 2 - \frac{1}{\alpha}$.

<u>2º Caso: $0 < \alpha < 1$:</u>

Temos a seguinte desigualdade:

$$\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha} \leqslant \left[\sum_{i=1}^{n} (\lambda p_i + \mu q_i)^{\alpha}\right]^{1/\alpha}$$

. Se $\alpha \dfrac{\beta-1}{\alpha-1} > 0$, isto é, $\beta < 1$,

$$\left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} \leqslant \left[\sum_{i=1}^{n} (\lambda p_i + \mu q_i)^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} . \tag{3.22}$$

Se além disso $\alpha \dfrac{\beta-1}{\alpha-1} < 1$, isto é, $\beta > 2 - \dfrac{1}{\alpha}$, a função $Y = x^{\left(\alpha\frac{\beta-1}{\alpha-1}\right)}$ é côncava donde:

$$\lambda\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} + \mu\left[\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} \leqslant \left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\alpha\frac{\beta-1}{\alpha-1}} \tag{3.23}$$

De (3.23) e (3.24) obtemos então:

$$\lambda\left[\sum_{i=1}^{n} p_i^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} + \mu\left[\sum_{i=1}^{n} q_i^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} \leqslant \left[\sum_{i=1}^{n} (\lambda p_i + \mu q_i)^{\alpha}\right]^{\frac{\beta-1}{\alpha-1}} \tag{3.24}$$

Subtraindo de (3.24) $\lambda+\mu=1$, e dividindo por $(2^{1-\beta}-1) > 0$, obtemos (3.20) e (3.21) para $0 < \alpha < 1$, e $2-\dfrac{1}{\alpha}<\beta<1$.

. Se $\alpha \frac{\beta-1}{\alpha-1} < 0$, isto é $\beta > 1$, a desigualdade (3.22) é invertida, o mesmo acontece com (3.23) visto que a função $Y = x^{\left(\alpha \frac{\beta-1}{\alpha-1}\right)}$ é convexa, consequentemente (3.24) também fica invertida; subtraindo desta $\lambda+\mu=1$, e dividindo por $(2^{1-\beta}-1) < 0$, obtemos novamente (3.20) e (3.21) para $0 < \alpha < 1$, $\beta > 1$.

3º Caso: $\beta = 2 - 1/\alpha$

Quando $\beta = 2 - \frac{1}{\alpha}$, temos o caso da informação da espécie $\gamma(\gamma=\frac{1}{\alpha})$ (a menos de uma constante multiplicativa) é estudado diretamente, utilizando a desigualdade (3.16).

Caso 3.1: $\gamma > 1$

$$\lambda \left[\sum_{i=1}^{n} p_i^{1/\gamma}\right]^{\gamma} + \mu \left[\sum_{i=1}^{n} q_i^{1/\gamma}\right]^{\gamma} \leqslant \left[\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{1/\gamma}\right]^{\gamma} \tag{3.25}$$

Subtraindo $\lambda+\mu=1$ à ambos os membros de (3.25), e dividindo uma nova desigualdade assim obtida por $(2^{\gamma-1}-1)>0,(\gamma>1)$, obtemos uma nova desigualdade:

$$\frac{\lambda}{2^{\gamma-1}-1} \left[\sum_{i=1}^{n} p_i^{1/\gamma}\right]^{\gamma} + \frac{\mu}{2^{\gamma-1}-1} \left[\sum_{i=1}^{n} q_i^{1/\gamma}\right]^{\gamma} \leqslant \frac{1}{2^{\gamma-1}-1} \left[\sum_{i=1}^{n} (\lambda p_i+\mu q_i)^{1/\gamma}\right]^{\gamma} \tag{3.26}$$

isto é,

$$\lambda\, {}_{\gamma}H(P) + \mu\, {}_{\gamma}H(Q) \leqslant {}_{\gamma}H(\lambda P+\mu Q) \tag{3.27}$$

Caso 3.2: $0 < \gamma < 1$

A desigualdade (3.25) é invertida, mas como $(2^{\gamma-1}-1)^{-1} < 0$ obtemos novamente (3.26) e consequentemente (3.27)

Podemos agora concluir que $H_\alpha^\beta(P)$ é côncava se:

. $\alpha > 1$ e $\beta \geqslant 2 - \frac{1}{\alpha}$

. $0 < \alpha < 1$ e $2 - \frac{1}{\alpha} \leqslant \beta < 1$

. $0 < \alpha < 1$ e $\beta > 1$.

Resumindo: Se $\alpha > 0$ e $\beta \geqslant 2 - \frac{1}{\alpha}$, $H_\alpha^\beta(P)$ é côncava ou seja:

Se $0 < \alpha \leqslant \frac{1}{2-\beta}$, H_α^β é côncava, desta forma provando o item (a).

Observação: Para o caso $\beta=1$, além de serem válidos os resultados de (a), outros resultados sobre a concavidade são obtidos, veja teorema 2.1.

b) Seja $P \in \Delta_n$, $\alpha > 0$ e $\beta=2$

$$H_\alpha^2(P) = -2\left[\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{1}{\alpha-1}} - 1\right]$$

1º Caso: $\alpha > 1$

Temos a seguinte desigualdade:

$$\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha} \geqslant \left[\sum_{i=1}^{n}(\lambda p_i + \mu q_i)^{\alpha}\right]^{1/\alpha}, \tag{3.28}$$

desde que $\frac{\alpha}{\alpha-1} > 0$ para $\alpha > 1$

$$\left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\frac{1}{\alpha-1}} \geqslant \left[\sum_{i=1}^{n}(\lambda p_i+\mu q_i)^{\alpha}\right]^{\frac{1}{\alpha}\cdot\frac{1}{\alpha-1}}. \tag{3.29}$$

Como para $\alpha > 1$, $\frac{\alpha}{\alpha-1} > 1$, a função $Y = x^{\frac{\alpha}{\alpha-1}}$ é convexa, donde:

$$\lambda\left[\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha}\right]^{\frac{\alpha}{\alpha-1}} + \mu\left[\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\frac{\alpha}{\alpha-1}} \geqslant$$

$$\geqslant \left[\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{1/\alpha} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{1/\alpha}\right]^{\frac{\alpha}{\alpha-1}}. \tag{3.30}$$

De (3.29) e (3.30) obtemos:

$$\lambda\left(\sum_{i=1}^{n} p_i^{\alpha}\right)^{\frac{1}{\alpha-1}} + \mu\left(\sum_{i=1}^{n} q_i^{\alpha}\right)^{\frac{1}{\alpha-1}} \geqslant \left[\sum_{i=1}^{n}(\lambda p_i+\mu q_i)^{\alpha}\right]^{\frac{1}{\alpha-1}}. \tag{3.31}$$

Subtraindo $1=\lambda+\mu$ à ambos os membros (3.31) e multiplicando a desigualdade resultante por -2 obtemos a nova desigualdade:

$$\lambda H_{\alpha}^{2}(P) + \mu H_{\alpha}^{2}(Q) = H_{\alpha}^{2}(\lambda P+\mu Q) \tag{3.32}$$

2º Caso: $0 < \alpha < 1$

Para $0 < \alpha < 1$ a desigualdade (3.18) fica invertida, desde que $\frac{\alpha}{\alpha-1} < 0$ $(0 < \alpha < 1)$ obtemos novamente (3.29) e visto que a função

$Y = x^{\frac{\alpha}{\alpha-1}}$ é convexa obtemos novamente (3.30) e (3.32) completando assim a demonstração de (b), e portanto do teorema.

TEOREMA 3.3: Para uma distribuição completa de duas probabilidades $(p,\ 1-p)$, $0 < \alpha < \frac{1}{2-\beta}$, a entropia paramétrica generalizada é côncava em $(\frac{1}{2},\ \frac{1}{2})$ e convexa em $(0,1)$ e $(1,0)$.

Prova:

Seja uma distribuição completa de duas probabilidades $(p,\ 1-p)$, a entropia paramétrica generalizada $H_\alpha^\beta(p,1-p)$ será estudada como função de $p \in (1,0]$. em $\mathbb{R}^+$.

$$H_\alpha^\beta(p,1-p) = \frac{1}{2^{1-\beta}-1}\left[(p^\alpha + (1-p)^\alpha)^{\frac{\beta-1}{\alpha-1}} - 1\right] .$$

$$\frac{d\ H_\alpha^\beta(p,1-p)}{dp} = \frac{\alpha(\beta-1)}{(2^{1-\beta}-1)(\alpha-1)}\left[p^\alpha+(1-p)^\alpha\right]^{\frac{\beta-1}{\alpha-1}-1}\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right]$$

$$\frac{d^2H_\alpha^\beta(p,1-p)}{dp^2} = \frac{\alpha(\beta-1)}{(2^{1-\beta}-1)(\alpha-1)} \left\{ \left[p^\alpha+(1-p)^\alpha\right]^{\frac{\beta-1}{\alpha-1}-1} (\alpha-1)\left[p^{\alpha-2}+(1-p)^{\alpha-2}\right] + \right.$$

$$\left. + \frac{\beta-\alpha}{\alpha-1}\left[p^\alpha+(1-p)^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \alpha\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right]\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right] \right\}$$

$$= \frac{(\beta-1)}{2^{1-\beta}-1}\,\alpha\left[p^\alpha+(1-p)^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \left\{\frac{\alpha(\beta-\alpha)}{(\alpha-1)^2}\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right]^2 + \right.$$

$$\left. + \left[p^\alpha+(1-p)^\alpha\right]\left[p^{\alpha-2}+(1-p)^{\alpha-2}\right] \right\} \tag{3.33}$$

Denotemos $C^+ = \dfrac{(1-\beta)}{2^{1-\beta}-1}\left[p^\alpha+(1-p)^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2}$ e

$$B = \frac{\alpha(\beta-\alpha)}{(\alpha-1)^2}\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right]^2 + \left[p^\alpha+(1-p)^\alpha\right]\left[p^{\alpha-2}+(1-p)^{\alpha-2}\right]$$

donde

$$\frac{d^2H_\alpha^\beta(p,1-p)}{dp^2} = -\,\alpha\, C^+\, B.$$

Seja $A = -\left[p^{\alpha-1}-(1-p)^{\alpha-1}\right]^2 + \left[p^{\alpha}+(1-p)^{\alpha}\right]\left[p^{\alpha-2}+(1-p)^{\alpha-2}\right]$

$$= \cancel{-p^{2\alpha-2}}+2p^{\alpha-1}(1-p)^{\alpha-1}\cancel{-(1-p)^{2\alpha-2}}\cancel{+p^{2\alpha-2}}+p^{\alpha}(1-p)^{\alpha-2}+(1-p)^{\alpha}p^{\alpha-2}\cancel{+(1-p)^{2\alpha-2}}$$

$$= p^{\alpha-2}(1-p)^{\alpha-2}\left[p^2 + 2p(1-p) + (1-p)^2\right]$$

$$= p^{\alpha-2}(1-p)^{\alpha-2}\left[p+(1-p)\right]^2 = p^{\alpha-2}(1-p)^{\alpha-2} .$$

portanto A é sempre positivo.

Se $\alpha \dfrac{(\beta-\alpha)}{(\alpha-1)^2} > -1$, então $B \geqslant A \geqslant 0$, o que implica que $H_\alpha^\beta(p,1-p)$ é côncava, pois $\dfrac{d^2H_\alpha^\beta(p,1-p)}{dp^2} \leqslant 0$.

Donde os resultados seguintes:

a) Se $0 < \alpha \leqslant \dfrac{1}{2-\beta}$ então $H_\alpha^\beta((p,1-p))$ é côncava.

b) Se $\alpha > 0$ e $\beta = 2$ então $H_\alpha^2((p,1-p))$ é côncava.

Confirmando assim os resultados do teorema 3.2 para uma distribuição completa de duas probabilidades (p,1-p).

Conforme (3.33) temos que:

$$\text{a)}\ \frac{d^2H_\alpha^\beta((1,0))}{dp^2} = \frac{d^2H_\alpha^\beta((0,1))}{dp^2} = \frac{\alpha(\beta-1)}{2^{1-\beta}-1}\left[0^\alpha+1^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2}\left\{\frac{\alpha(\beta-\alpha)}{(\alpha-1)^2}\left[0^{\alpha-1}-1^{\alpha-1}\right]^2 + \right.$$

$$\left. + \left[0^\alpha + 1^\alpha\right]\left[0^{\alpha-2} + 1^{\alpha-2}\right]\right\}$$

$$= \frac{\alpha(\beta-1)}{2^{1-\beta}-1}\left[\frac{\alpha(\beta-\alpha)}{(\alpha-1)^2} + 1\right] \tag{3.34}$$

Se $\frac{\alpha(\beta-\alpha)}{(\alpha-1)^2} < -1$, isto é,

para $0<\alpha>\frac{1}{2-\beta}$, temos: $\frac{d^2H_\alpha^\beta((0,1))}{dp^2} = \frac{d^2H_\alpha^\beta((1,0))}{dp^2} > 0$ (3.35)

$$\text{b)}\ \frac{d^2H_\alpha^\beta((\frac{1}{2},\frac{1}{2}))}{dp^2} = \frac{\alpha(\beta-1)}{2^{1-\beta}-1}\left[2(\tfrac{1}{2})^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2}\left\{\frac{\alpha(\beta-\alpha)}{(\alpha-1)^2}\left[(\tfrac{1}{2})^{\alpha-1}-(\tfrac{1}{2})^{\alpha-1}\right]+\left[2(\tfrac{1}{2})^\alpha\right]\left[2(\tfrac{1}{2})^{\alpha-2}\right]\right\} =$$

$$= \frac{\alpha(\beta-1)}{2^{1-\beta}-1}\left[(\tfrac{1}{2})^{\alpha-1}\right]^{\frac{\beta-2\alpha+1}{\alpha-1}}(\tfrac{1}{2})^{2\alpha-4} = \frac{\alpha(\beta-1)}{2^{1-\beta}-1}(\tfrac{1}{2})^{\beta-3}$$

Logo para $\alpha > 0$ temos:

$$\frac{d^2 H_\alpha^\beta((\frac{1}{2}, \frac{1}{2}))}{dp^2} = \frac{\alpha(\beta-1)}{2^{1-\beta}-1} \quad (\frac{1}{2})^{\beta-3} < 0 \tag{3.36}$$

Das relações (3.35) e (3.36) concluimos que para $0 < \alpha > \frac{1}{2-\beta}$ $H_\alpha^\beta((p,1-p))$ é côncava em $(\frac{1}{2}, \frac{1}{2})$ e convexa em (0,1) e (1,0); demonstrando assim o teorema. Colocando estes resultados graficamente temos:

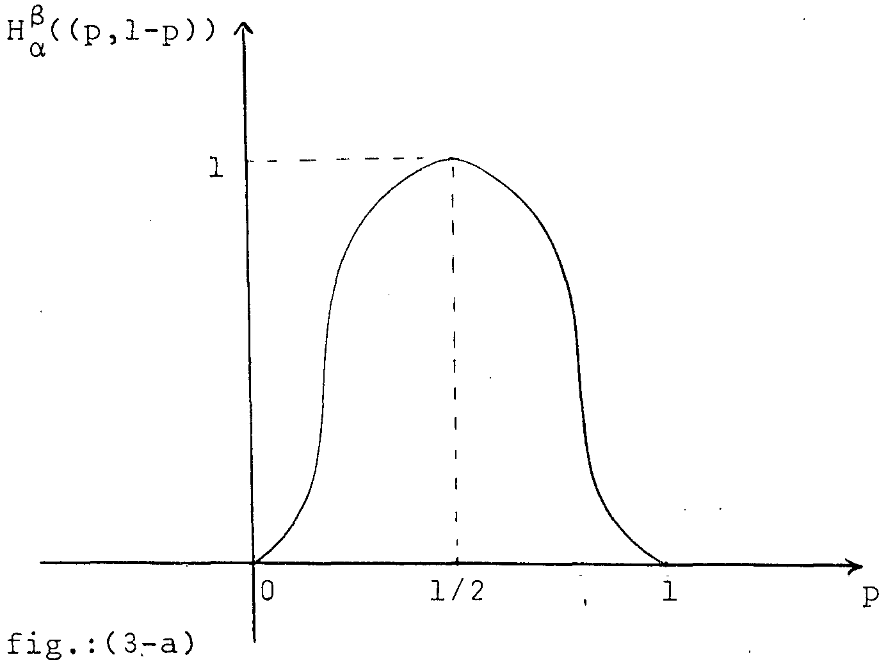


fig.:(3-a)

Fazendo q=1-p em (3.33) obtemos:

$$\frac{d^2 H_\alpha^\beta((p,q))}{dp^2} = \frac{\alpha(\beta-1)}{2^{1-\beta}-1} \cdot \left[p^\alpha + q^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \left\{ \frac{\alpha(\beta-\alpha)}{(\alpha-1)^2} \left[p^{\alpha-1} - q^{\alpha-1}\right]^2 + \right.$$

$$\left. + \left[p^\alpha + q^\alpha\right] \left[p^{\alpha-2} + q^{\alpha-2}\right] \right\}$$

Para $\beta \to 1$, e aplicando a regra de L'Hopital à expressão acima temos:

$$\lim_{\beta \to 1} \frac{d^2 H_\alpha^\beta((p,q))}{dp^2} = \lim_{\beta \to 1} \frac{-\alpha}{2^{1-\beta}\ell n\, 2} \left\{ \left[p^\alpha + q^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \left\{ \frac{\alpha(\beta-\alpha)}{(\alpha-1)^2} \left[p^{\alpha-1} - q^{\alpha-1}\right]^2 + \left[p^\alpha + q^\alpha\right]\left[p^{\alpha-2} + q^{\alpha-2}\right] \right\} + \right.$$

$$+ (\beta-1)\left[p^\alpha + q^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \ell n\left[p^\alpha + q^\alpha\right] \left\{ \frac{\alpha(\beta-\alpha)}{(\alpha-1)^2} \left[p^{\alpha-1} - q^{\alpha-1}\right]^2 + \left[p^\alpha + q^\alpha\right]\left[p^{\alpha-2} + q^{\alpha-2}\right] \right\} +$$

$$\left. + (\beta-1)\left[p^\alpha + q^\alpha\right]^{\frac{\beta-1}{\alpha-1}-2} \left\{ \frac{\alpha}{(\alpha-1)^2} \left[p^{\alpha-1} - q^{\alpha-1}\right]^2 \right\} \right\} =$$

$$= \frac{-\alpha}{\ell n 2} \left\{ \left[p^\alpha + q^\alpha\right]^{-2} \left\{ \frac{-\alpha}{\alpha-1} \left[p^{\alpha-1} - q^{\alpha-1}\right]^2 + \left[p^\alpha + q^\alpha\right]\left[p^{\alpha-2} + q^{\alpha-2}\right] \right\} \right\} =$$

$$= \frac{\alpha}{\ell n 2(1-\alpha)} \left[p^\alpha + q^\alpha\right]^{-2} \left\{ (\alpha-1)\left[p^\alpha + q^\alpha\right]\left[p^{\alpha-2} + q^{\alpha-2}\right] - \alpha\left[p^{\alpha-1} - q^{\alpha-1}\right]^2 \right\}.$$

Este resultado coincide com o resultado obtido no capítulo II para $\dfrac{d^2H_\alpha((p,1-p))}{dp^2}$, portanto para $\beta \to 1$ valem os resultados do teorema 2.1 item (b).

No que segue, o presente trabalho apresenta novas cotas superiores e inferiores sobre a probabilidade de erro em termos da entropia paramétrica generalizada; no que diz respeito à cotas inferiores será obtida uma desigualdade análoga à cota de Fano.

§ 3.2 - Cotas Sobre a Probabilidade de Erro em Termos de Entropia Paramétrica Generalizada

A - Cotas Superiores em Termos de Entropia Paramétrica Generalizada

Usando o fato que $H_\alpha^\beta(P)$ é uma função decrescente de α (lema 3.1) é obtida uma cota superior sobre a probabilidade de erro em termos de $H_\alpha^\beta(P)$.

TEOREMA 3.4: Para todo $n \geqslant 2$, $P \in \Delta_n$ e $\beta > 0$.

a) $1-p_M \leqslant \frac{1}{2} H_\alpha^\beta(P)$; para todo $\alpha > 0$, $\beta < 2$ se $\frac{1}{n} \leqslant p_M \leqslant \frac{1}{2}$,

b) Para $\beta < 2$, existe um $\alpha^*(P)$ tal que:

$$1-p_M > \frac{1}{2} H_\alpha^\beta(P) \quad \text{para todo} \quad \alpha > \alpha^*(P) \text{ se } \frac{1}{2} < p_M \leqslant 1$$

c) $1-p_M \leqslant \frac{1}{2} H_\alpha^\beta(P)$; para todo $\alpha > 0$, $\beta > 2$ se $\frac{1}{2} < p_M \leqslant 1$

d) Para $\beta > 2$ existe um $\alpha^*(P)$ tal que:

$$1-p_M \geqslant \frac{1}{2} H_\alpha^\beta(P) \text{ para todo } \alpha > \alpha^*(P) \text{ se } \frac{1}{n} \leqslant p_M \leqslant \frac{1}{2}$$

e) Para $\beta = 2$, $1-p_M \leqslant \frac{1}{2} H_\alpha^2(P)$; $\forall\, P \in \Delta_n$.

Prova:

Seja $\beta > 0$.

Pelo lema 3.2 (d) temos que:

$$\lim_{\alpha\to\infty} H_\alpha^\beta(P) = H_\infty^\beta(P) = (2^{1-\beta}-1)^{-1}\left[p_M^{\beta-1} - 1\right] \text{ para todo } P \in \Delta_n.$$

Consideremos a função:

$$f(p_M) = 1-p_M - \frac{1}{2}(2^{1-\beta}-1)^{-1}\left[p_M^{\beta-1} - 1\right] \tag{3.37}$$

com:

$$f'(p_M) = -1 - \frac{1}{2}(2^{1-\beta}-1)^{-1}(\beta-1)\, p_M^{\beta-2} \tag{3.38}$$

e

$$f''(p_M) = \frac{-1}{2}(\beta-1)\,(\beta-2)\,(2^{1-\beta}-1)^{-1}\, p_M^{\beta-3} \tag{3.39}$$

Notemos que:

$$f(\tfrac{1}{2}) = f(1) = 0 \tag{3.40}$$

. Se $0 < \beta < 2$, $f''(p_M) < 0$ para todo $p_M \varepsilon (0,1]$ isto indica que $f(p_M)$ é uma função côncava e atinge seu único máximo em $f'(p_M) = 0$, isto é,

$$\frac{-1}{2}(2^{1-\beta}-1)^{-1}(\beta-1)\, p^{\beta-2} = 1 \Longrightarrow p_M = \left[\frac{-2(2^{1-\beta}-1)}{\beta-1}\right]^{\frac{1}{\beta-2}}, \tag{3.41}$$

portanto os únicos zeros de $f(p_M)$ são os indicados em (3.40), ou seja, $p_M = \frac{1}{2}$ e $p_M = 1$, o que nos leva aos seguintes resultados para $0 < \beta < 2$:

$$f(p_M) \leqslant 0, \quad \text{para } \frac{1}{n} \leqslant p_M \leqslant \frac{1}{2} \tag{3.42}$$

$$f(p_M) \geqslant 0, \quad \text{para } \frac{1}{2} < p_M \leqslant 1 \tag{3.43}$$

De (3.42) e (3.43) respectivamente obtemos:

$$\text{para } \frac{1}{n} \leqslant p_M \leqslant \frac{1}{2}\,;\quad 1-p_M \leqslant \frac{1}{2}(2^{1-\beta}-1)^{-1}\left[p_M^{\beta-1} - 1\right] =$$

$$= \frac{1}{2}\lim_{\alpha\to\infty} H_\alpha^\beta(P) = \frac{1}{2} H_\infty^\beta(P) \tag{3.44}$$

$$\text{para } \frac{1}{2} < p_M \leqslant 1;\ 1 - p_M \geqslant \frac{1}{2}(2^{1-\beta}-1)^{-1}\left[p_M^{\beta-1} - 1\right] =$$

$$= \frac{1}{2}\lim_{\alpha\to\infty} H_\alpha^\beta(P) = \frac{1}{2} H_\infty^\beta(P) \tag{3.45}$$

Desde que $H_\alpha^\beta(P)$ é uma função decrescente de α para todo β fixo deduzimos (a) de (3.44).

De fato:

$$1-p_M \leqslant \frac{1}{2} H_\alpha^\beta(P) \leqslant \frac{1}{2} H_\alpha^\beta(P), \text{ para todo } \alpha > 0 \text{ e } \frac{1}{n} \leqslant p_M \leqslant \frac{1}{2} .$$

Para demonstrar o item (b) consideremos:

$$\alpha^*(P) = \inf_\alpha \ \alpha / 1-p_M > \frac{1}{2} H_\alpha^\beta(P) . \tag{3.46}$$

Desde que $H_\alpha^\beta(P)$ é uma função estritamente decrescente de α para todo $P \neq P^o = (\frac{1}{n},\ldots,\frac{1}{n})$, deduzimos (b) de (3.45).

De fato:

$$1-p_M \geqslant H_\infty^\beta(P) = H_{\alpha^*(P)}^\beta(P) > H_\alpha^\beta(P), \text{ para todo } \alpha>\alpha^*(P) \text{ e } \frac{1}{2} < p_M \leqslant 1$$

. Analogamente à (a) e (b) demonstramos (c) e (d) respectivamente, partindo do fato que para $\beta > 2$, $f''(p_M) > 0$, isto é, $f(p_M)$ é convexa. Também o item (e) é demonstrado analogamente à (a) partindo do fato que para $\beta = 2$ $f(p_M) = 0$, $\forall\ 0 \leqslant p_M \leqslant 1$.

<u>Observação</u>: A quantidade $\alpha^*(P)$ definida em (3.46) depende também de n, mas não consideraremos o fato afim de simplificar a notação.

Os seguintes teorema e lema provam que:

a) $\alpha^* = \dfrac{1}{2-\beta}$ se $0 < \beta < 2$ e $\beta \neq 1$

b) $\alpha^* = 0$ se $\beta \geqslant 2$

onde:

$$\alpha^* = \inf_P \{\alpha^*(P)/P \in \Delta_n, \quad n \geqslant 2\} \tag{3.47}$$

<u>TEOREMA 3.5</u>: Para todo $P \in \Delta_n$, $0 < \alpha \leqslant \dfrac{1}{2-\beta}$, $n \geqslant 2$,

$$1-p_M \leqslant \frac{1}{2} H_\alpha^\beta(P) \tag{3.48}$$

Prova:

Provaremos o teorema para n=2 e extenderemos então à $n > 2$.

Para o caso n=2, o lema 3.2 garante que:

$$\frac{1}{2} H_\alpha^\beta((0,1)) = \frac{1}{2} H_\alpha^\beta((1,0)) = 0 \quad \text{e,}$$

$$\frac{1}{2} H_\alpha^\beta((\tfrac{1}{2}, \tfrac{1}{2})) = \frac{1}{2}$$

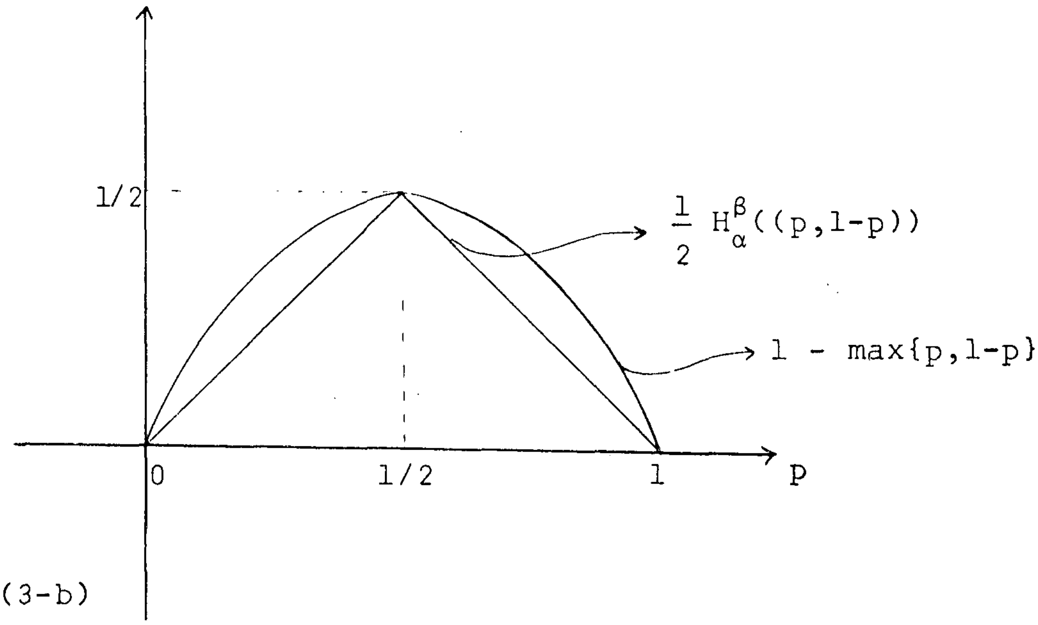


fig.: (3-b)

Desde que o gráfico de 1-max{p,1-p}, $0 \leq p \leq 1$, consiste de duas retas entre (0,0) e $(\frac{1}{2}, \frac{1}{2})$ e entre $(\frac{1}{2}, \frac{1}{2})$ e (1,0) ver figura (3.b), obtemos o resultado desejado da concavidade de $H_\alpha^\beta((p,1-p))$, $0 < \alpha \leq \frac{1}{2-\beta}$, e $0 < p \leq 1$ ou seja:

$$1-p_M \leq \frac{1}{2}\, H_\alpha^\beta(P), \qquad p \in \Delta_2.$$

Seja $P \in \Delta_n$, $n > 2$, sem perda de generalidade, assumimos que $p_M = p_n$. Pelo lema 3.2 (e) temos que:

$$H_\alpha^\beta(1-p_M,p_M) \leq H^\beta(P), \text{ para todo } P \in \Delta_n,\ \beta > 0{,}2 > 0 \tag{3.49}$$

$$\text{onde } p_M = \max_i\, p_i,$$

desde que o teorema é válido para n=2, temos:

$$1-p_M \leq \frac{1}{2}\, H_\alpha^\beta(1-p_M,\, p_M), \text{ para todo } P \in \Delta_n \tag{3.50}$$

$$p_M \in (0,1] \quad e \quad 0 < \alpha \leq \frac{1}{2-\beta}$$

De (3.49) e (3.50) deduzimos (3.48) e assim completamos a demonstração do teorema.

Para o caso $\beta=2$ temos do teorema 3.2 (b) que $H_\alpha^\beta(P)$ é uma função côncava com respeito à P para todo $P \in \Delta_n$ e $\alpha > 0$, assim, analogamente ao raciocínio acima concluímos novamente o teorema para $\beta = 2$.

<u>LEMA 3.3</u>: Para todo $0 < \alpha > \frac{1}{2-\beta}$, $\beta > 0$, $\beta \neq 1$, $\beta \neq 2$ existe um conjunto não vazio $K \subset \Delta_n$ tal que:

$$1-p_M > \frac{1}{2} H_\alpha^\beta(P), \text{ para todo } P \in K \tag{3.51}$$

<u>Observação 1</u>: Se $\beta > 2$ o lema é válido para todo $\alpha > 0$.

<u>Observação 2</u>: Pelo teorema 3.4 temos que:

a) Se $\beta < 2$, todo $P \in K$ é tal que $p_M > \frac{1}{2}$.

b) Se $\beta > 2$, todo $P \in K$ é tal que $p_M \leq \frac{1}{2}$.

Prova:

É suficiente provar o lema para $n = 2$, desde que o caso $n > 2$ pode ser obtido do caso $n=2$ adicionando zeros.

Dado $0 < \alpha > \frac{1}{2-\beta}$, denotemos p' e p" pontos onde $\frac{1}{2} H_\alpha^\beta(p,1-p)$ muda de convexa para côncava e de côncava para convexa. Do teorema 3.2 concluimos que $0 < p' < \frac{1}{2}$ e $\frac{1}{2} < p'' < 1$.

Seja

$$K = \{(p,1-p)/0 < p < p' \quad \text{ou} \quad p'' < p < 1\} \tag{3.52}$$

Portanto, para todo $P \varepsilon K$

$$1-p_M > \frac{1}{2} H_\alpha^\beta(P).$$

Isto completa a demonstração do lema.

<u>Observação 3</u>: As quantidades p' e p" não são necessariamente os limites superior e inferior respectivamente do conjunto de pontos onde (3.52) é válida.

Notemos que o teorema 3.5 vale em particular para C/x, qualquer que seja x, de modo que:

$$1-\max\{P(1/x),\ldots,P(n/x)\} \leqslant \frac{1}{2} H_\alpha^\beta(C/x), \tag{3.53}$$

$$\text{para todo} \quad 0 < \alpha \leqslant \frac{1}{2-\beta}.$$

Tomando expectativas com respeito a x de ambos os lados de (3.53) obtemos:

$$P(e) \leqslant \frac{1}{2} I_\alpha^\beta, \qquad \text{para todo} \qquad 0 < \alpha \leqslant \frac{1}{2-\beta}. \tag{3.54}$$

De fato:

$$P(e/x) = 1-\max\{P(1/x),\ldots,P(n/x)\}$$

$$E_x\ P(e/x) = E_x\left[1-\max\{P(1/x),\ldots,P(n/x)\}\right]$$

$$P(e) = 1 - \max\{P(1/x),\ldots,P(n/x)\}.$$

Além disso:

$$E_x\left[H_\alpha^\beta(C/x)\right] = I_\alpha^\beta, \text{ para } \alpha,\beta > 0.$$

Se supormos que para todos os valores possíveis de x, $p(e/x) \geqslant 1/2$, obtemos em consequência do teorema 3.4 uma cota mais geral:

$$P(e) \leqslant \frac{1}{2} I_\alpha^\beta \;, \quad \text{para} \quad 0 < \beta \leqslant 2 \;\text{ e }\; \alpha > 0 \tag{3.55}$$

O lado direito de (3.54) e (3.55) é chamado cota I_α^β. Segue do lema 2.3 que a extensão de (3.53) para $0 < \alpha > \frac{1}{2-\beta}$ e todo $P \in \Delta_n$ é inválida.

B - Cotas Inferiores em Termos de Entropia Paramétrica Generalizada

Nesta seção cotas inferiores sobre a probabilidade de erro derivadas da entropia paramétrica generalizada são desenvolvidas por analogia à cota de Fano.

Definimos

$$\bar{h}_\alpha^\beta(p_e) = \sup_P \{H_\alpha^\beta(P)/P \in \Delta_n,\ 1\text{-}\max\{p_1,\ldots,p_n\} = p_e\} \tag{3.56}$$

TEOREMA 3.6: Para $\alpha,\beta > 0$, $\alpha \neq 1$, $\beta \neq 1$, o máximo de $\bar{h}_\alpha^\beta(p_e)$ é atingido quando todos os p_i=1,...,n (exceto talvez um) são iguais à $p_e/(n-1)$ e então

$$\bar{h}_\alpha^\beta(p_e) = \left[2^{1-\beta}-1\right]^{-1}\left\{\left[(1-p_e)^\alpha + \left(\frac{1}{n-1}\right)^{\alpha-1} p_e^\alpha\right]^{\frac{\beta-1}{\alpha-1}} - 1\right\} \qquad (3.56)$$

Prova:

Sejam $\alpha,\beta > 0$, $\alpha \neq 1$, $\beta \neq 1$, $P \in \Delta_n$ com

$$p_e = 1 - \max_i p_i.$$

No teorema 2.6 foi demonstrado que

$$H_\alpha^\beta(P) \leq \frac{1}{1-\alpha} \log \left[(1-p_e)^\alpha + \left(\frac{1}{n-1}\right)^{\alpha-1} p_e^\alpha\right] \qquad (3.57)$$

com a igualdade se e somente se $p_i = \frac{p_e}{n-1}$, $i=1,\ldots,n$ e $p_n = 1-p_e$. Desde que

$$H_\alpha^\beta(P) = g_\beta(H_\alpha(P)) \quad \text{onde}$$

$g_\beta(x) = (2^{1-\beta}-1)\left[2^{(1-\beta)x} - 1\right]$, é uma função crescente de x, temos:

$$H_\alpha^\beta(P) = g(H_\alpha(P)) \leq g\left[\frac{1}{1-\alpha} \log \left[(1-p_e)^\alpha + \left(\frac{1}{n-1}\right)^{\alpha-1} p_e^\alpha\right]\right]$$

$$= (2^{1-\beta}-1)^{-1}\left\{2^{(1-\beta)\frac{1}{1-\alpha}\log\left[(1-p_e)^\alpha+\left(\frac{1}{n-1}\right)^{\alpha-1} p_e^\alpha\right]} - 1\right\}$$

$$= (2^{1-\beta}-1)^{-1}\left\{\left[(1-p_e)^\alpha+\left(\frac{1}{n-1}\right)^{\alpha-1} p_e^\alpha\right]^{\frac{\beta-1}{\alpha-1}} - 1\right\} \qquad (3.58)$$

Tomando o supremo em ambos os lados da desigualdade acima temos (3.56) e portanto completamos a demonstração do teorema.

TEOREMA 3.7: Para $\alpha > 0$, $\beta \geqslant 2 - \frac{1}{\alpha}$, temos uma generalização da cota de Fano.

$$H_\alpha^\beta(C/x) \leqslant (2^{1-\beta}-1)^{-1}\left\{\left[(1-P_e)^\alpha + (n-1)\left(\frac{P_e}{n-1}\right)^\alpha\right]^{\frac{\beta-1}{\alpha-1}} - 1\right\} \quad (3.59)$$

Prova:

Substituindo p_i por $P(C/x_i)$, para cada $X = x_i$ em (3.58), onde $P_e = 1-Ex(P(C/x))$. Além disso, tomando a expectativa com respeito à X de ambos os lados de (3.58) obtemos:

$$H_\alpha^\beta(C/x) = (2^{1-\beta}-1)^{-1}\left\{E_x\left[(1-p_e)^\alpha + (n-1)^\alpha\left(\frac{p_e}{n-1}\right)^\alpha\right]^{\frac{\beta-1}{\alpha-1}} - 1\right\} \quad (3.60)$$

Da expressão (3.19) e desde que (3.25) fica invertida para $0 < \gamma = \frac{1}{2} < 1$, temos:

$$\lambda\left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}} + \mu\left(\sum_{i=1}^{n} q_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}} \geqslant \left[\sum_{i=1}^{n}(\lambda p_i^\alpha + \mu q_i^\alpha)^\alpha\right]^{\frac{\beta-1}{\alpha-1}} \quad (3.61)$$

onde λ e μ ε $[0,1]$, $\lambda + \mu = 1$ para todo $\alpha > 1$ e $\beta \geqslant 2 - \frac{1}{\alpha}$.

É fácil verificar que (3.61) ainda é válida se $0 < \alpha < 1$ e $\beta > 1$.

Isto implica que a função $f(P) = \left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}}$ é convexa se $\alpha > 1$ e $\beta \geq 2 - \frac{1}{\alpha}$, ou se $0 < \alpha < 1$ e $\beta > 1$.

A desigualdade (3.61) fica invertida se $0 < \alpha < 1$ e $2 - \frac{1}{\alpha} \leq \beta < 1$ [veja as desigualdades (3.24) e (3.25)].

Isto implica que a função $f(P) = \left(\sum_{i=1}^{n} p_i^\alpha\right)^{\frac{\beta-1}{\alpha-1}}$ é côncava para todo $0 < \alpha < 1$ e $2 - \frac{1}{\alpha} \leq \beta < 1$.

Portanto pela desigualdade de Jensen, temos:

$$E_x\left[(1-p_e)^\alpha + (n-1)\left(\frac{p_e}{n-1}\right)^\alpha\right]^{\frac{\beta-1}{\alpha-1}} \gtreqless \left\{\left[E_x(1-p_e)\right]^\alpha + (n-1)\left[E_x\left(\frac{p_e}{n-1}\right)\right]^\alpha\right\}^{\frac{\beta-1}{\alpha-1}}, \qquad (3.62)$$

De acordo com $\begin{cases} \alpha > 1, & \beta \geq 2 - \frac{1}{\alpha} \\ 0 < \alpha < 1, & \beta > 1 \end{cases}$ ou

$$0 < \alpha < 1, \; 2 - \frac{1}{\alpha} \leq \beta < 1$$

Mas:

$$(2^{1-\beta}-1) \lesseqgtr 0 \text{ de acordo com } \beta \gtreqless 1. \qquad (3.63)$$

Das relações (3.60), (3.62), obtemos

$$H_\alpha^\beta(C/x) \leq (2^{1-\beta}-1)^{-1}\left\{\left[\left[E_x(1-p_e)\right]^\alpha + (n-1)\left[E_x\,\frac{p_e}{n-1}\right]^\alpha\right]^{\frac{\alpha-1}{\alpha-1}} - 1\right\} \text{ ou}$$

$$H_{\alpha}^{\beta}(C/x) \leq (2^{1-\beta}-1)^{-1} \left\{ \left[(1-P_e)^{\alpha} + (n-1)\left(\frac{P_e}{n-1}\right)^{\alpha} \right]^{\frac{\beta-1}{\alpha-1}} - 1 \right\} \quad \text{para todo } \alpha > 0, \; \beta \geq 2 - \frac{1}{2} \; .$$

Isto completa a demonstração do teorema proposto.

CASOS PARTICULARES:

(i) Quando $\alpha = \beta$, $\beta > 0$, (3.59) reduz-se à:

$$H^{\beta}(C/x) \leq \left[(1-P_e)^{\beta} + (n-1)\left(\frac{P_e}{n-1}\right)^{\beta} - 1 \right] (2^{1-\beta}-1)^{-1} \; ,$$

a qual é a cota inferior estudada por Devijver [8] (1977) para o caso da entropia de grau β.

(ii) Quando $\gamma = \frac{1}{2} = 2 - \beta$, portanto $\gamma > 0$ ($\alpha > 0$, $\beta \geq 2 - \frac{1}{\alpha}$)

(3.59) reduz-se à:

$$_{\gamma}H(C/x) \leq \left\{ \left[(1-P_e)^{1/\gamma} + (n-1)\left(\frac{P_e}{n-1}\right)^{1/\gamma} \right]^{\gamma} - 1 \right\} (2^{\gamma-1}-1)^{-1} \quad \text{ou}$$

$$_{R}H(C/x) \leq \left\{ \left[(1-P_e)^{R} + (n-1)\left(\frac{P_e}{n-1}\right)^{R} \right]^{\frac{1}{R}} - 1 \right\} \frac{R}{R-1}$$

para todo $R > 0$, (quando $R = \frac{1}{\alpha}$) a qual é a cota inferior para a medida de informação R-NORMA estudada por Boekee e Lubbe [4] (1980).

(iii) Quando $\beta \to 1$ temos $\lim_{\beta \to 1} H_\alpha^\beta(C/x) = H_\alpha(C/x)$, que é a entropia condicional de ordem α. Neste caso (3.59) reduz-se à:

$$H_\alpha(C/x) \leqslant (1-\alpha)^{-1} \log \left\{ (1-P_e)^\alpha + (n-1)\left(\frac{P_e}{n-1}\right)^\alpha \right\}, \text{ para todo } \alpha > 0$$

que é a mesma cota inferior estudada por Toussaint [28] (1977).

<u>Observação</u>:

a) Toussaint [28] (1977) em seu trabalho prova o resultado somente para o caso $\alpha > 1$, mas o resultado acima prova que também é válido para $\alpha \leqslant 1$, isto é, para todo α positivo. O caso separado para $0 < \alpha < 1$ foi estudado por Taneja [23].

b) O resultado obtido em (i) vale para todo $\beta > 0$, porém Devijver [8] (1977) já havia provado aquele resultado para $\beta > 1$. O caso separado $0 < \beta < 1$ foi analisado por Taneja [24].

(iv) Para o caso em que $\beta \to 1$, $\alpha \to 1$ temos:

$$\lim_{\substack{\alpha \to 1 \\ \beta \to 1}} H_\alpha^\beta(C/X) = H(C/X)$$

a qual é a entropia condicional de Shannon.

Neste caso de (3.59) obtém-se:

$$H(C/x) \leqslant -(1-P_e)\log(1-p_e)-P_e\log\frac{P_e}{n-1}$$

que é a bem conhecida cota de Fano.

Notemos que tomando $\alpha = 2$ na relação (3.58) obtemos:

$$H_2^\beta(P) \leqslant (2^{1-\beta}-1)^{-1}\left\{\left[(1-p_e)^2 + \left(\frac{1}{n-1}\right)^{2-1} p_e^2\right]^{\frac{\beta-1}{2-1}} - 1\right\} \tag{3.65}$$

$$\beta > 0,\ \beta \neq 1$$

ou seja:

$$(2^{1-\beta}-1)^{-1}\left\{\left[\sum_{i=1}^{n} p_i^2\right]^{\beta-1} - 1\right\} \leqslant (2^{1-\beta}-1)^{-1}\left\{\left[(1-p_e)^2+\left(\frac{1}{n-1}\right)p_e^2\right]^{\beta-1} -1\right\}, \tag{3.66}$$

$$\beta > 0,\ \beta \neq 1$$

Para $\beta > 1$, (3.66) implica que:

$$\sum_{i=1}^{n} p_i^2 \geqslant (1-p_e)^2 + \left(\frac{1}{n-1}\right)p_e^2$$

que é a desigualdade (2.55) obtida no capítulo II a qual explicita a seguinte solução para p_e

$$\frac{n}{n-1}\left\{1 - \sqrt{\frac{n \sum_{i=1}^{n} p_i^2 - 1}{n-1}}\right\} \leqslant p_e \tag{3.67}$$

Substituindo p_i por $P(C/x_i)$, para cada $X = x_i$ em (3.67), onde $P_e = 1-E_x(P(C/x))$, e tomando a expectativa com respeito a X de ambos os lados da desigualdade (3.66) obtemos:

$$E_x\left\{\frac{n}{n-1}\left[1-\sqrt{\frac{n\left(\sum_{i=1}^{n} P^2(C/x_i)\right)-1}{n-1}}\right]\right\} \leq P_e \qquad (3.68)$$

Aplicando a desigualdade de Jensen à função côncava $\sqrt{\ }$ obtemos

$$\frac{n-1}{n}\left[1-\sqrt{\frac{n\,E_x\left[\sum_{i=1}^{n} P^2(C/x)-1\right]}{n-1}}\right] \leq P_e \qquad (3.69)$$

Claramente (3.68) é mais próxima da igualdade que (3.69).

Estes resultados já foram comentados no capítulo II.

Da relação (3.59) teorema 3.7 , temos que a solução explícita de P_e em termos de $H_\alpha^\beta(C/X)$ não é possível para quaisquer valores de α e β, contudo para $\alpha = 2$ e $\beta > 1$ obtemos:

$$P_e \geqslant \frac{n-1}{n}\left\{1-\sqrt{\frac{n\,|\,(1+(2^{1-\beta}-1)\,H_2^\beta(C/x)^{\frac{1}{\beta-1}}}{n-1}}\right\} \qquad (3.70)$$

De fato:

Tomando $\alpha = 2$ na relação (3.59) temos:

$$H_2^\beta(P) \leqslant (2^{1-\beta}-1)^{-1}\left\{\left[(1-P_e)^2 + \frac{P_e^2}{n-1}\right]^{\beta-1} - 1\right\} \qquad (3.71)$$

Para $\beta > 1$, (3.71) implica que:

$$(2^{1-\beta}-1)\, H_2^\beta(P) \geqslant \left[(1-P_e)^2 + \frac{P_e^2}{n-1}\right]^{\beta-1} - 1$$

$$1 + (2^{1-\beta}-1)\, H_2^\beta(P) \geqslant \left[(1-P_e)^2 + \frac{P_e^2}{n-1}\right]^{\beta-1}$$

$$\left[1 + (2^{1-\beta}-1) H_2^\beta(P)\right]^{\frac{1}{\beta-1}} \geqslant 1-2P_e + P_e^2 + \frac{P_e^2}{n-1}$$

$$\left[1 + (2^{1-\beta}-1)\, H_2^\beta(P)\right]^{\frac{1}{\beta-1}} \geqslant 1-2P_e+\left(1 + \frac{1}{n-1}\right)P_e^2$$

$$n\left[1+(2^{1-\beta}-1)\, H_2^\beta(P)\right]^{\frac{1}{\beta-1}} \geqslant n-2nP_e + \frac{n^2}{n-1}\, P_e^2$$

$$n\left[1+(2^{1-\beta}-1)\, H_2^\beta(P)\right]^{\frac{1}{\beta-1}} - 1 \geqslant (n-1) - 2nP_e + \frac{n^2}{n-1}\, P_e^2$$

$$\frac{1}{n-1}\left\{n\left[1+(2^{1-\beta}-1)\, H_2^\beta(P)\right]^{\frac{1}{\beta-1}} - 1\right\} \geqslant 1 - 2\,\frac{n}{n-1}\, P_e + \frac{n^2}{n-1}\, P_e^2$$

$$\frac{1}{n-1}\left\{n\left[1+(2^{1-\beta}-1)\ H_2^{\beta}(P)\right]^{\frac{1}{\beta-1}} - 1\right\} \geqslant \left[1 - \frac{n}{n-1}\, P_e\right]^2$$

$$\sqrt{\frac{n\left[1+(2^{1-\beta}-1)\ H_2^{\beta}(P)\right]^{\frac{1}{\beta-1}} - 1}{n-1}} \geqslant 1 - \frac{n}{n-1}\, P_e$$

$$1 - \sqrt{\frac{n\left[1+(2^{1-\beta}-1)\ H_2^{\beta}(P)\right]^{\frac{1}{\beta-1}} - 1}{n-1}} \leqslant \frac{n}{n-1}\, P_e$$

$$\frac{n-1}{n}\left\{1 - \sqrt{\frac{n\left[1+(2^{1-\beta}-1)\ H_2^{\beta}(P)\right]^{\frac{1}{\beta-1}} - 1}{n-1}}\right\} < P_e\ .$$

Fazendo $\beta = \frac{3}{2}$ obtemos:

$$P_e \geqslant \frac{n-1}{n}\left\{1 - \sqrt{\frac{n\left[1+(2^{-1/2}-1)\ \frac{1}{2}\ H(P)\right]^2 - 1}{n-1}}\right\} \tag{3.72}$$

A qual é a explicitação de P_e em termos da medida de informação R-Norma (para $r = \frac{1}{1/2} = 2$) obtida por Boeke e Lubbe [4] (1980).

BIBLIOGRAFIA

[1] ACZEL, J. and DARÓCZY, Z.: On Measures of Information and Their Characterization, Academic Press, New York, 1975, Ch.5.

[2] ARIMOTO, S.: "Information - Theoretical Considerations on Estimation Problems, Information and Control, Vol. 19, nº 3, October 1971, pag. 181-194.

[3] BEN-BASSAT, M. and RAVIV, J.: Rényi's Entropy and the Probability of Error, IEEE Transactions Theory, Vol. II-24, nº 3, May 1978, pag. 324-331.

[4] BOEKEE, D.E. and VAN DE LUBBE, J.C.A.: The R-Norm Information Measure, Information and Control, Vol. 45, 1980, pg. 136-155.

[5] COVER, T.M. and HART, P.E.: Nearest Neighbor Pattern Classification, IEEE Transaction Information Theory, Vol. IT-13, 1967, pag. 21-27.

[6] DARÓCZY, Z.: Generalized Information Functions, Information and Control, 16, 1970, pg. 36-51.

[7] DEVIJVER, P.A.: On a New Class of Bounds on Bayes Risk in Multihypothesis Pattern Recognition, IEEE Trans. Comput., Vol. C-23, nº 1, 1973, pag. 70-80.

[8] DEVIJVER, P.A.: Entropies of Degree β and Lower Bounds for the Average Error Rate, Information and Control, Vol. 34, 1977, pag. 222-226.

[9] FERGUSON, T.S.: Mathematical Statistics, Academic Press, New York, 1967, pag. 291-297.

[10] GALLAGER, R.G.: Information Theory and Reliable Communication, New York, Wiley, 1968, pag. 520-521.

[11] HARDY, G.H., LITTLEWOOD, J.E. and POLYA, G.: Inequalities, Cambridge University Press, 1967.

[12] HAVDRA, J. and CHARVAT, F.: Quantification Method of Classification Process The Concept of Strutural a-Entropy, Kybernetika 3, 1967, pag. 30-35.

[13] HELLMAN, M.E. and RAVIV, J.: Probability of Error, Equivication and Chernoff Bound, IEEE Transactions on Information Theory, Vol. IT-16, 1970, pag. 368-372.

[14] KOVALESKY, V.A.: "The Problem of Caracter Recognition from the Point of View of the Mathematical Statistics", in Character Readers and Pattern Recognition, V.A. Kovalesky, Ed New York, Spartan, 1968.

[15] MANGASARIAN, O.L.: Pseudo-Convex Functions, Sian J. Control, Vol. 3, 1965, pag. 281-290.

[16] MANGASARIAN, O.L.: Nonlinear Programming, McGraw-Hill, New York, 1969.

[17] MATHAI, A.M. and RATHIE, P.N.: Basic Concepts in Information Theory and Statistics, Wiley Eastern Limited, New Delhi, 1975.

[18] RENYI, A.: On Measures of Entropy and Information, Proceedings of the 4^{th} Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, Vol. I, 1961, pag. 547-561.

[19] RENYI, A.: Probability Theory, Amsterdan: North Holland, 1970.

[20] SHANNON, C.E.: A Mathematical Theory of Communication, Bell System Tech. J. 27, 1948, pag. 379-423 e 623-656.

[21] SHARMA, B.D. and MITTAL, D.P.: New Non-additive Measures of Entropy for Discrete Probability Distributions, Journal of Mathematical Sciences, 10, 1975, pag. 28-40.

[22] TANEJA, I.J.: A Joint Characterization of Shannon's Entropy and Entropy of Type β Througn a Functional Equation - J. Math. Sci., Vol. 10, 1975, pag. 69-74.

[23] TANEJA, I.J.: Some Contributions to Information Theory - I. (A Survey). On Measures of Information, J. Comb. Inform. & System. Sci., Vol. 4, 1979, pag. 259-280.

[24] TANEJA, I.J.: Comentário sobre o paper "A Generalization of Shannon's Equivocation and the Fano Bound" - Comunicada em 1981.

[25] TANEJA, I.J.: Comentário sobre o paper "Entropies of Degree β and Lower Bounds for the Average Error Rate", Comunicada em 1981.

[26] TANEJA, I.J.: Generalized n-Entropy and Probability of Error, Comunicada.

[27] TOUSSAINT, G.T.: On Information Transmission, Nonparametric Classification and Measuring Depéndence Between Randon Variables, Proc. Symp. Stattistic and Related Topics, Carleton Univ. Ottawa, Canada, 1974, pag. 30-1 to 30-9.

[28] TOUSSAINT, G.T.: A Generalization of Shannon's Equivocation and the Fano Bound, IEEE Transactions on Systems, and Cybernetics, 1977, pag. 300-302.

[29] VAN DER PYL, T.: Information d'Ordre α e de Type β Axiomatique et Propriétis, Thèse de 3º Cycle, Paris (1977).